DOCUMENT RESUME

ED 181 053	TM 009 957
TITLE	Proceedings of the Invitational Conference on Testing Problems (22nd, New York, New York, November 1, 1958).
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	1 Nov 58
NOTE	130p.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Cognitive Measurement; *College Entrance Examinations; Conference Reports; Early Childhood Education; Educational Practice; *Effective Teaching; Elementary Secondary Education; Foreign Countries; Higher Education; *Maladjustment; Personality Assessment; Predictive Measurement; Response Style (Tests); *Scholarships; Self Concept; *Testing Problems; Testing Programs; Test Selection
IDENTIFIERS	USSR

ABSTRACT

The conference was planned to appeal to a diverse audience and was not centered on a single unifying theme. It was divided into three sessions. Session I was comprised of two presentations which focused on measurement problems at the preschool, early childhood, and preadolescent levels: Measurement of Cognitive Abilities at the Preschool and Early Childhood Level, by Dorothea A. McCarthy; and Prediction of Maladjustive Behavior, by William C. Kvaraceus. Two papers were presented at Session II, exemplifying basic and applied research: A Theory of Test Response, by Jane Loevinger; and Measurement and Prediction of Teacher Effectiveness, by David G. Ryans. Session III was composed of a panel which discussed: What Kinds of Tests for College Admission and Scholarship Programs, by Robert L. Ebel and by Alexander G. Wesman; Criteria for Selecting Tests for College Admissions and Scholarship Programs, by John C. Flanagan; and The Nature of the Problem of Improving Scholarship and College Entrance Examinations, by E. F. Lindquist. The luncheon address was delivered by Henry Chauncey and was entitled, Some Observations on Soviet Education. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181053

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Proceedings

1958

Invitational Conference on Testing Problems

TM009957

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational
Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EDUCATIONAL TESTING SERVICE

Copyright, 1959, Educational Testing Service
20 Nassau Street, Princeton, N. J.
Library of Congress Catalog Number: 47-11220
Printed in the United States of America

Invitational Conference on Testing Problems

November 1, 1958

Hotel Roosevelt, New York City

ROGER T. LENNON, Chairman

EDUCATIONAL TESTING SERVICE

Princeton, New Jersey Los Angeles, California

4

ETS Board of Trustees, 1958-59

Henry H. Hill, *Chairman*

Arthur S. Adams
Frank D. Ashburn
George F. Baughman
Grace V. Bird
Frank H. Bowles
John T. Caldwell
James B. Conant
John W. Gardner
Caryl P. Haskins
Philip J. Hickey
Frederick L. Hovde
Wallace Macgregor
Lester Nelson
O. Meredith Wilson

Officers

Henry Chauncey, *President*
William W. Turnbull, *Executive Vice President*
Henry S. Dyer, *Vice President*
Robert L. Ebel, *Vice President*
Catherine G. Sharp, *Secretary*
G. Dykeman Sterling, *Treasurer*
William H. Reuter, *Assistant Treasurer*

Foreword

At the twenty-second annual Invitational Conference on Testing Problems, some 600 participants considered perhaps as wide a variety of topics as has ever been covered at one of these meetings. The presentations ranged over different facets of measurement from preschool age through adulthood, from research in psychological theory to methodology in attacking the problem of delinquency, from questions on the measurement of teacher effectiveness to questions about the kinds of tests used in college admissions and scholarship programs. Taken as a whole, the papers reflected the tremendous scope of the measurement field and its relation to some of the most pressing educational and social questions of our time.

To fashion such diversified fare into an interesting, cohesive program is no small challenge, and we are all indebted to Roger T. Lennon, Chairman of the 1958 Invitational Conference, for meeting that challenge so successfully. Dr. Lennon, Director of the World Book Company's Division of Test Research and Service, is a man who has been influential in guiding the development of the testing movement along the path of a responsible and progressive approach to the entire field of measurement. We are indeed grateful to him for drawing on his knowledge of the field to produce such a stimulating program.

We are equally indebted to the principal speakers and discussants for presenting such "food for thought." I have no doubt but what the ideas and suggestions for future study presented at this Conference will be much discussed in educational and measurement circles from now on, and will, eventually, lead to fresh, productive approaches to both old and new problems.

HENRY CHAUNCEY
President

Preface

The annual Invitational Conference on Testing Problems has achieved a reputation as perhaps the outstanding meeting of the year for leaders in the testing field. With each passing year the consistent excellence of the programs has attracted a steadily growing audience. As the number of those attending has increased, so also has the range of their special interests within the measurement field broadened; and the task of developing a program of high appeal to the diversified interests represented in the audience has become increasingly demanding on the Chairman.

Theorist and practitioner, test producer and test consumer, personality researcher and guidance counselor, state testing director and university instructor—all bring their varied needs and interests and backgrounds to the Conference, each confidently expecting a program that will be uniquely appealing and satisfying to *him;* and rarely have they been disappointed. The 1958 Conference was planned to be of broad and diverse appeal, to match in range of content the heterogeneity of the audience—even to the abandoning of the tradition of a unifying "theme" for the meeting.

The opening session of the 1958 Conference comprised two papers which focused attention on measurement problems at the preschool, early childhood, and preadolescent level—an area which has received perhaps somewhat less than its proper share of research attention in recent years. Developments and problems in the measurement of cognitive abilities at this level were clearly and comprehensively reviewed by Dr. Dorothea McCarthy. Dr. William Kvaraceus spoke on the measurement and prediction of maladjustive behavior, presenting a penetrating analysis of the conceptual and methodological problems in the appraisal of certain personal and social characteristics.

In the second session, papers by Dr. Jane Loevinger and Dr. David G. Ryans represented neat exemplifications of basic and applied research, the former directed to the elaboration of a new construct in personality measurement, and the latter to a review of certain findings of the massive Teacher Characteristics Study which Dr. Ryans has been directing for several years. Dr. Loevinger presented a closely reasoned argument for the existence of a fundamental personality variable underlying responses to many personality measures. Dr. Ryans not only presented major substantive findings of the Teacher Characteristics Study with respect to measurement and prediction of teacher effectiveness, but also provided a critique of methodology in this area. Dr. David Tiedeman presented a brief discussion of the Loevinger thesis, and Dr. Harry Gilbert invited attention to some of the practical implications of Dr. Ryans' work for problems of teacher recruitment, selection, and merit rating.

At the luncheon meeting Dr. Henry Chauncey presented an absorbing account of his observations about Soviet education, based on his tour of Soviet Russia in early 1958. Dr. Chauncey's report that educational and psychological measurement as we know it is virtually non-existent in the schools of Russia is but one illustration of the many provocative elements in his description of an educational system so different from our own in philosophy and practice, and yet so full of significance for our way of life.

The afternoon session brought together Drs. Robert L. Ebel, John C. Flanagan, E. F. Lindquist, and Alexander G. Wesman in a panel discussion on the topic "What Kinds of Tests for College Admissions?" These authorities, constituting perhaps as expert a panel on this problem as might be assembled, surprised the audience with their diverse, not to say conflicting, views. Their lively presentations illumined an area which is assuming ever greater importance in our education.

It is a pleasure for me to record here my great appreciation to the participants in the 1958 Conference for their uniformly fine contributions, and to Educational Testing Service which played the role of host organization with its accustomed graciousness and efficiency. I trust that I express the consensus of the 600 or more leaders in measurement who attended the Conference in voicing the belief that the 1958 Conference did not fall short of the high standards set in previous meetings.

ROGER T. LENNON
Chairman

Contents

Foreword by **Henry Chauncey** 3
Preface by **Roger T. Lennon** 5

Session I

Remarks of the Chairman, **Roger T. Lennon**, Director, Division of Test Research and Service, World Book Company 9
Measurement of Cognitive Abilities at the Preschool and Early Childhood Level, **Dorothea A. McCarthy**, Fordham University. 10
Prediction of Maladjustive Behavior, **William C. Kvaraceus**, Boston University and Director Juvenile Delinquency Project, National Education Association 26

Session II

Remarks of the Chairman 35
A Theory of Test Response, **Jane Loevinger**, Jewish Hospital of St. Louis 36
Discussion, **David V. Tiedeman**, Harvard University 47
Comment, **Jane Loevinger** 51
Measurement and Prediction of Teacher Effectiveness, **David G. Ryans**, University of Texas 52
Discussion, **Harry B. Gilbert**, New York City Board of Education.. 67

Luncheon Address

Some Observations on Soviet Education, **Henry Chauncey**, Educational Testing Service 71

Session III

Remarks of the Chairman 87

What Kinds of Tests for College Admission and Scholarship Programs? ROBERT L. EBEL, Educational Testing Service 88

Criteria for Selecting Tests for College Admissions and Scholarship Programs. JOHN C. FLANAGAN, University of Pittsburgh and the American Institute for Research 98

The Nature of the Problem of Improving Scholarship and College Entrance Examinations, E. F. LINDQUIST, State University of Iowa 104

What Kinds of Tests for College Admission and Scholarship Programs? ALEXANDER G. WESMAN, The Psychological Corporation 114

Appendix 121

Session I

Remarks of the Chairman

ROGER T. LENNON, Director, Division of Test Research and Service, World Book Company

I should like to open this first session of the 1958 Invitational Conference on Testing Problems by voicing sentiments that I am sure you all share with me, sentiments of appreciation to Educational Testing Service, our host organization, for its continuing courtesy in making possible these meetings that have, over the years, proven so stimulating to all of us in the field of measurement.

At this first session today, we shall consider two different topics, each touching on measurement problems in areas where research has lagged behind the need for sound measurement techniques. The first area is the measurement of cognitive abilities in very young children. I can think of no better person to review for us what has been going on in that area, and to point out the measurement problems and difficulties there, than Dorothea McCarthy. Her work, and her special research in the field of language development, is familiar to all who are interested in child development. Therefore, it is with a great deal of pleasure that I will present to you the first speaker, Dr. Dorothea A. McCarthy, of Fordham University, talking on "Measurement of Cognitive Abilities at the Preschool and Early Childhood Level."

Immediately following Dr. McCarthy's presentation, we shall move on to consider measurement questions related to the social problem that we call juvenile delinquency. Here is an area where, generally speaking, measurement people have been conspicuous by their absence rather than their contribution. An outstanding exception to this generalization is the man who will be our second speaker, a man who has pioneered in introducing good measurement and research techniques into the work of preventing and controlling delinquent behavior in young people. He is Dr. William C. Kvaraceus, of Boston University, who is currently on a one-year leave from that institution to direct the National Education Association's Juvenile Delinquency Project. He will talk to us this morning about the "Prediction of Maladjustive Behavior."

11

Measurement of Cognitive Abilities at the Preschool and Early Childhood Level

DOROTHEA A. McCARTHY, Professor of Psychology, Graduate School of Arts and Sciences, Fordham University

Even the casual observer of a group of young children of preschool age is immediately impressed with marked individual differences in the behavior of children of the same age and with the very marked contrasts in the behavior of children only six months or a year apart in chronological age. In motor performances, which it is possible to observe most objectively, older children are able to do things faster, more smoothly and with greater ease and precision than younger children. They also manifest greater strength and are able to attempt and perform more complex tasks. These things hold true whether the observer is concerned with gross motor performances or with fine muscular coordination.

Turning to the other broad area of observable behavior, the verbal or linguistic, through which the child is able to give some reflection of his concept formation and the higher thought processes, it is clear that older children talk more, know more words, and put them together in longer and more complex groups than do the younger children. In their handling of words and numbers, older children show their ability to be more clear and specific in contrast to the vagueness which characterizes the expression of younger children. Also evident is the increasing ability of older children to handle abstract ideas, in contrast to the concrete ideas which are typical of younger children. The degree of complexity of the abstract ideas they are able to use increases with age, and they also show increasing mental alertness and speed in their ability to solve problems which most people would agree are of an intellectual or cognitive nature.

These are some of the qualities of the effectiveness of mental functioning which the lay person refers to when he says that a young child "has a good head on his shoulders," that "he will go far," or that he is "bright for his age." The marked changes which occur over a span of a few months, and the sharp contrasts in the mental performances of different children of the same age, serve to highlight the enormity of the individual differences the psychologist attempts to measure. These are especially noteworthy during the preschool period, for it has often been shown that individual differences in any function are always most marked during periods of rapid development. All these evidences of mental growth are occurring in extremely complex organisms who are growing physically in varying environmental circumstances, and adapting to varying numbers and kinds of persons, both children and adults. These processes of mental growth are also occurring in children who are clinging or independent, withdrawing or aggressive, patient or flighty, and so on for an infinite variety of personality traits which are more or less closely related to aspects of mental development which we have made the most successful attempts to measure.

Merely because we can think of intellectual functions in the abstract and try to measure them in isolation does not mean that they occur in isolation in nature, any more than the chemist who isolates iron in the laboratory finds it in pure form in nature. Just as most chemical elements are found in compounds of varying degrees of complexity, so, too, children's minds are developing in youngsters who are receiving a wealth of warm, affectionate nurturance; in those who are lonely, deprived and neglected; in those who are thwarted, punished and rejected at every turn; as well as in those who are frightened and shocked into complete withdrawal by their early infantile experiences.

Binet (the father of the mental testing movement, on whose ideas we are still elaborating) was well aware of most of these things, and he gave us a tool which, when adapted in this country by Terman's genius, has proven to be our best yardstick for children of school age for a whole generation. This tool, in spite of its many limitations and shortcomings and current obsolescence, pointed the way for the now widely accepted group testing movement at all educational levels.

While psychologists and statisticians have been pushing mass testing and finding better, more reliable methods of testing large groups with multiple-choice items which can be entered on answer sheets and scored by machine for easy and efficient reporting to school administrators, little progress has been made in developing tools for younger age levels which still need individual examination.

There is at the present time a tremendous need for new tools which will do at the infant and preschool levels for today's generation what the Stanford-Binet did for the last generation of children at school age. The increased birthrate since World War II has produced a bumper crop of preschool children known as the "baby boom," so there is a large per-

13

centage of the population now in the age bracket which requires individual mental examination. Preventive work and early diagnosis of, and planning for, appropriate education of handicapped children creates a demand for good and even highly specialized tools. There are more handicapped children who survive nowadays than in former times, due to improved obstetrics and pediatrics. The great wave of interest in mentally retarded children, due to the banding together of their parents to share their common problems, and the appropriation of huge sums of money for the study, care and education of the mentally retarded has caught the psychological profession short-handed with only outmoded tools. It is as if we were trying to plough a field with a horse instead of a tractor.

As knowledge about what constitutes normal child development has spread, deviations from what is considered normal are bringing children to the attention of child guidance clinics and other service agencies at earlier ages. Unfortunately, too, our culture has produced increased numbers of illegitimate children in need of placement for adoption at early ages. Many childless couples are eager to adopt these children, but they have come to expect some sort of assurance of normality in the children they propose to adopt, and hope for reasonable accuracy of the predictions that are made. Since so many studies have shown the value of early placement in a family setting and the harmful effects on mental growth of life in an institution, and mental tests have not proven very effective in infancy, placement agencies have recently been encouraging adoption even without benefit of tests.

There are several reasons which seem to have accounted for the neglect of this area in the field of psychometrics. The training institutions have been preoccupied with training clinicians for the Veterans Administration and for work with adults in mental hospitals, and give their trainees only a minimum of opportunity to work with children. Child development research centers have become preoccupied with longitudinal follow-up studies which are just now beginning to emerge. After the disappointing results of some of the long-term prediction studies, interest has turned, among research workers, to deeper and more penetrating studies of children's thinking and reasoning; to studies of their concepts of causal relationships, of space, time, and number. Usually these processes are studied in a small sample of available children rather than in representative samples. Such qualitative studies, many of them suggested by the very provocative work of Piaget, have not aimed to develop practical tests or measuring instruments. However, they should prove highly suggestive to the constructors of tests of the future. For example, Harrison (18) reports a correlation of .7 between mental age and children's comprehension of words involving time concepts.

The group-testing movement with the resultant handling of mass data has taught us much about norms, the importance of large numbers, the techniques of stratified sampling, item analysis, weighting and reliability, so that our standards as to what constitute adequate norms have advanced remarkably and have become so high that it is no longer feasible for an individual to develop a test which will withstand the critical scrutiny of his colleagues and gain widespread acceptance without substantial financial assistance.

Usually mental testing is done by women who generally find it easier than men to relate to young children and to establish rapport with them. Many men seem to think it a threat to their masculinity to work with young children, and with their greater interest in things mathematical and statistical, and their greater economic responsibilities, they are usually diverted to the field of group testing which yields greater financial rewards. Women, on the other hand, often marry and do not remain in the field, or, if they do continue active professionally, they usually are in service agencies or in teaching rather than in research settings where they might have opportunities to develop new tools.

Young children are especially baffling and thwarting to the scientific research workers who must have infinite patience in order to work with them. They sleep a large portion of the working day and woe-betide the investigator whose work conflicts with a preschool child's regular nap-time! Then, too, young children suffer from many contagious diseases which frequently occasion broken appointments and loss of time. Preschool children cannot come for examinations alone and must always be accompanied to a testing center by a mother or other adult who may have other responsibilities and often finds it difficult to cooperate.

Even when such practical difficulties are overcome, preschool children are notorious for their shyness and negativism at certain ages, so that the test administrator, after investing an appreciable amount of time in a case, may find he has an incomplete test because of refused items. In addition, preschool children further frustrate examiners with their fleeting attention spans and their vivid imaginations which often confuse fact and fantasy. Furthermore, they cannot read or write or even mark an answer sheet for machine scoring. So, it is little wonder they have been neglected for easier and more lucrative fields.

There is also the ever-present ticklish problem of validity. What constitutes intelligent behavior from age two to age five? Usually we have been content with tests that show developmental changes with chronological age. While this is a necessary and important characteristic of a

15

test for young children, many other things, like height and weight, also show increments with age and obviously are not the kinds of things we wish to measure when we are studying cognitive processes. In the preschool field we do not even have age grade progress or school grades or teachers' estimates of intelligence or achievement test scores as criteria against which to check our instruments, and much further work needs to be done along these lines. As Landreth (24) points out, "no attempt has so far been made to relate representative popular judgments of behavior capacities of young children to their performance on 'intelligence' tests." (p. 333)

Many predictive studies have been conducted over the years and although there have been minor variations due to different sampling methods, the use of different tests, different examiners, and different intervals over which the prediction has been attempted, the predictive value of all the tests seems to be inversely proportional to the age at which the first test is given and to the interval between tests. Maurer (28) states: "the studies that have been made show a disappointing lack of positive correlation between early standing and later standing when the intervals between tests are long enough for such information to be useful." (p. 20) Summarizing the literature, in 1949, Bayley (4) said: "The results of these studies are interesting but have not so far given us any adequately predictive batteries of tests," and with regard to her own data she said: "In all of the comparisons so far made on the Berkeley Growth Study children, little consistency in relative scores could be found during the first two to four years. After this age, however, intellectual progress becomes fairly stable." (p. 168)

The most recent of the longitudinal investigations is a comprehensive study from the Fels Research Institute by Sontag, Baker and Nelson (30) involving retests on 140 children with the Stanford Binet scales from ages three to twelve. The authors conclude: "The data descriptive of the IQ performance of the entire group was much like data of other comparable longitudinal studies. . . . The pattern of retest correlation of IQ's at one age with IQ's at later ages was similar to the pattern found in other studies in the literature. Correlations decreased as the age interval between the two tests was lengthened and increased as the child grew older, if the interval between the two tests was held constant. However, the inter-age correlations in the preschool period were slightly higher than those previously reported in the literature, suggesting the possibility that the correlations during the preschool period may have been somewhat underestimated in the past." These authors used a technique of smoothed trend lines to minimize the effect of errors of measurement found in any one test. They report that "the preschool tests were found to be only slightly more variable about this trend line than the tests administered during the elementary school years." (p. 52)

Confirming what Bayley (4) has called "lability" of test scores, the Fels study (30) reports on the "idiosyncratic nature of the patterns of change" which were found. "It would appear that the extent of IQ change found during childhood has been previously underestimated. (The median amount of change in this instance was 17.9 IQ points.) Sixty-two per cent of the children changed more than 15 IQ points sometime during the course of mental development from the age of 3 to age 10" (pp. 53-54). Another interesting finding was that certain children seem to have accelerative or decelerative tendencies to their mental growth curves that appear to be quite independent of special abilities in types of tests passed.

Because of the failure of long-term prediction of *infant* tests, many people have tended to discard all *preschool* tests. This seems to be quite unfortunate, for there *are* tests which have good reliability and do give reasonably accurate predictions after about two years of age. There is a marked change in the testability of children after the onset of speech and Goodenough and Maurer (14) found that verbal and nonverbal abilities are readily differentiable on a fairly permanent basis as early as three years of age. Landreth (24) cites Honzik's (21) study to the effect that by four years of age children's test scores correlate to the extent of about .6 with the scores they earn at six years of age. Thus, it is possible in nursery school to predict fairly accurately which children will be ready for first grade at age six, and to aid in their school placement well enough to avoid some serious misplacements and perhaps to avoid many of the tragic cases of reading disability which emerge from the early grades and graduate to our reformatories before adolescence.

So far, the best tools for the measurement of mental functioning have been the so-called tests of general intelligence which measure only abstract verbal ability. The reason we consider them "best" is because they do a fairly satisfactory job, better than any other tools do, of predicting academic success. We have compulsory education laws which force children into academic situations which traditionally have used highly verbal techniques. To succeed in school, children must achieve early mastery of the language arts of listening, speaking, reading and writing, as well as spelling and some facility in dealing with numbers. These so-called tool subjects must be mastered if the child is to be able to study content subjects and learn something of his cultural heritage. This situation has tended to make us keep a narrow focus in our intelligence testing. One of the main things which prevents correlations from going much higher is that many children who have abilities along nonverbal lines do not have opportunity to show them in the traditional school or to reflect them in achievement test scores or other criteria. As soon as such individuals are released from the pressures of a verbal academic situation and have opportunities to learn trades and earn their way in the practical school of experience they do much better than we would have predicted on the basis of their early scores on verbal tests.

Intelligence, as we use the term in psychometrics, is, after all, an abstract term. We can keep it narrow in focus if we wish to make predictions to a fairly narrow or specific criterion such as academic success. However, Thorndike (34) long ago pointed out that abstract intelligence was only one aspect and that there were other types of intelligence which he called concrete intelligence, or the ability to deal with things (perhaps our performance tests are getting at this), and also social intelligence, or the ability to deal with people, which we have largely ignored in the field of measurement.

As our knowledge of child development becomes broader and richer and deeper, however, we can broaden our concept of intelligence and try to predict in other areas of ability than the strictly verbal. Gesell pointed the way in his developmental schedules which recognized the four dimensions of locomotor, adaptive, linguistic and personal-social behavior on an intuitive basis, but he lost the precision of his careful observations and control of conditions in his overall subjective appraisal of the developmental quotient and the lack of statistical equivalence of his various scales.

Tests yielding a profile of subscores would be most helpful for purposes of differential diagnosis. These can be developed either on an intuitive basis or by means of factor analysis. Thurstone (35) has developed the latter type most fully in his Primary Mental Abilities Scale, but this is a group test, and at the earliest age level with which we are concerned the sub-tests do not possess satisfactory reliability.

Ruth Griffiths (15) in her volume *Abilities of Babies* described, in 1954, an interesting scale for the first two years of life which yields a profile of five subscores as well as an overall General Quotient. The subscales developed on an intuitive basis are the Locomotor Scale, the Personal-Social Scale, the Hearing and Speech Scale, one for Eye and Hand Development, and a Performance Scale. She presents strikingly different profiles showing the relatively low performance of deaf babies on the Hearing and Speech scale, of blind children on the Eye-Hand scale, and of spastic infants on the Locomotor scale.

This test is based on 571 children in England whose fathers' occupations correspond roughly to the distribution of the Minnesota Scale of occupations developed by Goodenough for the 1940 U.S. Census data. It is not known, however, how representative they are of British children or how British and American children whose fathers have the same occupations compare on intelligence. This scale has not as yet been widely used in this country. The author is extending it upward into the preschool and early childhood range. It seems that the greatest contribution it makes as an infant scale is the introduction of a Speech and Hearing scale, an area which previous infant tests have largely ignored or sampled very sparingly.

As the writer has pointed out elsewhere (26), most scales at the infant level have been largely sensori-motor in character, probably because these aspects of behavior are in the ascendancy at the earliest levels and are most readily observable without instrumentation or highly specialized training. They have, for the most part, ignored any attempt to study the development of the precursors of the verbal factor in the early prelinguistic babblings of the infant. The remarkable studies of Irwin (22), however, have shown quite clearly that phonemic analyses of these early utterances do yield developmental trends which in many ways behave in much the same way as do our verbal tests at higher age levels. They differentiate between children living in family and in institution settings, even in the first six months of life. One of my students, Regina Fisichelli (11), largely confirmed Irwin's findings on 100 infants between 6 and 18 mos. of age. Subsequently, when Catalano and I (9) followed up approximately one third of her group at an average age of 41 months with the Stanford-Binet Form L, we found correlations of .5 between certain measures of infant vocalizations such as consonant vowel ratio and later Stanford-Binet IQ. These substantial positive correlations were obtained with a sample which represented only the lower end of the distribution and were maintained even when ages at time of both tests were held constant.

Although our initial attempts to measure mental ability at the infant level did not prove as fruitful as we had hoped, it does not mean that it cannot be done with improved techniques. In this area of measurement we are in a stage analogous to the period when James McKeen Cattell, who was the first to use the term mental test, was discouraged because his attempts to predict the academic achievement of Columbia students with a battery involving strength of grip and speed of color naming tests and the like proved to be a disappointment. We must try again with improved techniques.

This is being done by two of the best qualified and most careful research workers I know. Nancy Bayley, who developed the California First Year Mental Scale and who has done more longitudinal follow-up work than any one else, is now working on a new infant scale. Dr. Katharine Maurer Cobb has recently returned from South Africa, where she examined large numbers of primitive infants, and is now gathering data on an American sample. I have high hopes that these new scales being developed by experienced research workers will give us much improved tools for the next generation of babies.

As for the preschool and early childhood levels, I myself am working on a new battery of tests which I hope will eventuate in a new point scale with a profile of several subscores appraising various aspects of the child's development. A preliminary pilot study on approximately 100 mentally retarded children between the ages of 6 and 14 years yielded good age trends and differentiated well between educable and non-educable institutionalized children. Further data on normal children are now being analyzed and it is hoped that before too long a full-scale standardization with stratified sampling will begin with the items which are most promising, through the gracious assistance of the Psychological Corporation.

One of the major problems with mental tests is the problem of obsolescence. This matter of becoming outdated affects not only materials, but also content and norms. Most examiners are aware, when using the Stanford-Binet, of the obsolescent upright telephones, the black stoves with top ovens, the high laced shoes and the steam locomotives which many preschool children have never seen. But perhaps we do not stop to realize that in this day of miracle drugs people rarely die from the flu, and besides we now call it a virus. Terman and Merrill could not have known that sulfa would be discovered the year their 1937 revision was published and thus invalidate one of their favorite verbal absurdities. In this day of numerous telephones and automobiles it is indeed rare to see a uniformed messenger boy delivering a telegram by bicycle, and with the advent of automatic dryers the sight of clothes drying on a line, as occurs in two Stanford-Binet pictures, is far less familiar than it used to be.

I suspect, also, that our norms are quite obsolete, for today's children are being exposed to a much more varied and stimulating environment than the children of twenty years ago. The invention of plastics has made the manufacture of cheap toys in greater variety possible. And the higher standard of living means that more children possess a greater variety of toys. Witness, too, the growth of toy stores and the nursery furniture industry.

The fact that children travel in cars throughout infancy and preschool life means their geographical orientation must be greatly expanded over that of children of a generation ago who scarcely left their own yards or their own block until school age. The rapid spread of nursery schools has also stimulated more preschool children and probably accelerated their mental and social growth.

Just as anthropometric studies have shown that children tend to be taller and heavier than they were a generation ago, so psychological norms need revision and updating from time to time. The recent monograph by Templin (32) which gathered data on children's language development, using the same recording techniques and the same sampling techniques in the same city where I worked twenty-five years before, showed children using on the average one more word per sentence, equal to almost a year's acceleration, and showing corresponding advances in other aspects of language development. Undoubtedly any verbal test would reflect similar up-grading in children's performance, due to the influence of radio, television, better standards of living, more leisure time which parents spend with children, less use of illiterate nursemaids, less bilingualism and greater permissiveness in dealing with children of today. It is probable that obsolescence of materials which makes tests more difficult for children counteracts the effect of their greater facility in language development in the total test score, so that we are lulled into complacent continued acceptance of our old instruments by an artificial stability in mean score for groups.

Another tendency which I think is unfortunate for the whole field of measurement is that with the development of highly specialized clinics for various types of cases in rehabilitation centers, speech and hearing clinics and cerebral palsy centers, workers not highly trained in psychometrics are being forced to develop their own batteries of tests and are using groups of items which are not standardized and on which they develop their own subjective norms on the basis of their clinical experience. We in the field of measurement and psychometrics probably would not like to call these instruments tests at all, but they are being used to give diagnoses on large numbers of cases in the service agencies because we have lagged behind and have not supplied the kinds of tools which are needed for the urgent problems of differential diagnosis. Ingenious scales of this sort have been developed by De Hirsch (10) for children with language disorders and by Haeussermann (17) for cerebral palsied children. The problem is: what do normal children do with the same tasks? Anyone who works for a period of time with one type of handicapped child is bound to develop a distorted norm biased in favor of such cases and while discriminations of degree of defect may be made carefully, the perspective in relation to the normal child will be lost without the use of standardized tests based on representative sampling.

Baldwin (3) gives a provocative discussion of the problem of determining what are the primary mental abilities to be tested. Much of the confusion in the literature stems from the fact that certain preconceptions as to the nature of intelligence must go into the preliminary selection of items included in any battery of tests. No matter how elaborate the subsequent statistical treatment of scores is, we can never get out of a factor analysis elements that were not present in the battery of tests to begin with. Factor analysis may point out a few interrelationships which we were not astute enough to anticipate. Correlations obtained between two or more factors may be due not to any real relationship existing between the two abilities, but to the fact that the two tests are measuring or involve certain common elements in the environment. Tests may appear to be related to each other because both are related to a third element which may or may not be an ability. It appears that Hofstaetter's (20) analysis of Bayley's 1933 data contributed little that had not been deduced earlier from a careful study of the standard deviations and the correlations at successive ages in relation to the content of the test.

With the many practical difficulties of locating subjects for individual testing, mentioned earlier, one wonders if the necessary preliminary testing in order to form a correlational matrix *before* setting up a new battery will prove feasible. It can work out well in group testing at higher levels, as in the work of Guilford in personality testing, but I doubt whether individual tests for young children will be established on the basis of previously determined empirical factors for some time to come. Factor analysis can always be performed *after* a battery is prepared intuitively, and duplicating or irrelevant material can subsequently be discarded. If we dare to try out new ideas we may find out things which can be confirmed by statistical analysis but which we might be much longer in discovering by purely empirical methods.

Because there has been so little success in measuring a variety of abilities in infants, the hypothesis has been advanced that the various mental abilities differentiate with advance in age. The evidence for this hypothesis is not entirely clear. It may be that rudiments of several abilities are present early, but that our techniques have been too crude thus far to isolate them and to measure them for study in infancy. Irwin's (22) contribution on infant language is a striking example of how refined techniques have opened up an area for investigation in infancy which was completely overlooked before.

From the early days of testing, when we cherished the naive hope that mental tests might somehow measure native capacity apart from environmental influences, we have come through the era in which we have admitted that we always measure hereditary factors *and* environmental influences interacting, and that all we ever get in an intelligence test is present functional level. We have seen numerous studies which have attempted to estimate how much each of these factors contributes, and the most recent studies are trying to determine *how* and *why* certain environmental influences operate the way they do.

There are many parallels and areas of overlap between the development of language and the measurement of intelligence, for, as Baldwin (3) states, "In the development of conceptual thought, language plays a very significant role. . . . The word as a sign of an object implies his (the child's) ability to maintain some sort of mental representation of the object, action or situation that the word signifies . . . (and) the word for an abstraction is very convenient because it gives the abstraction a concrete handle, the concept is more easily used because there is a word for it." (p. 354) There should be little wonder, then, that our best mental tests are verbal and that vocabulary tests have proven their usefulness time and again.

Research on language development in children is therefore at the very core of measurement of cognitive functions. Earlier investigations revealed that children's language development varied with paternal occupation and that children talked in a more advanced fashion the more contact they had with adults. In fact, there is considerable evidence which seems to indicate that the more intense and the more prolonged the contact with the mother, the more accelerated the language development.

The amazing findings of Brodbeck and Irwin (8) indicate that the environmental impact on language development is measurable even in the first three months of life, where infant speech sounds are much more advanced for children raised in a normal family setting than for institution infants. Wyatt (37) has spelled out rather clearly how the mechanism of unconscious identification is "the common denominator of many of the behavioral events of interaction between mother and child and, in particular, for the facts of mutual imitation, so essential for the learning of language."

In this connection the work of Goldfarb (13) is of particular interest, for it seems to point out some of the dynamics involved in the acquisition of language as the groundwork for conceptual thought which we try to measure with our mental tests at higher ages. Goldfarb, who compared institutionalized children and those who grew up in normal family settings, points out that in the family the child is cared for by "*specific* adults called parents" who are warm and loving and who minister to the child with "detailed understanding," and that such contact is continuous and affords constant stimulation from the same source. Hence it is unambiguous and more readily comprehensible than fleeting attention from a variety of nurses in an institution. Goldfarb says: "he (the child) receives active encouragement, for example, to babble, make sounds and then words. . . . Finally, the child's relationship to his parents includes a strong element of *reciprocation*." Thus, he sees the family as a source of tender emotions which provides the setting necessary for transfer of functions from the parent to the child in the process of identification. The child who thus experiences love, sympathy and affection in the cradle learns to trust those about him; he learns to wait, and to delay immediate satisfactions, and hence develops inner control, planfulness and foresight, which Goldfarb claims is the basis for conceptual thought and cognitive development. Klatskin (23) at Yale has shown that infants raised on a flexible rooming-in arrangement tested considerably higher on the Cattell Infant Scale than the normative children of the Harvard Growth Study who were subjected to rigid schedules with less mothering.

Children who do not enjoy warmth of affection in family settings are psychologically deprived. It is probably impossible to separate the intellectual from the affective aspects in these early experiences in infancy, but we probably never succeed completely in so doing even at higher ages, for intelligence is only one aspect of the total behavior of the organism and cannot be measured entirely in isolation. The clinical literature is replete with evidences of intelligence test scores which are depressed when children are suffering from anxiety and are raised when children are experiencing periods of relative calm. While, of course, some children can go through crises of adjustment apparently without having their disturbances reflected in test scores, life experiences do seem related to lability of test scores with sufficient frequency to deserve much more serious consideration than they usually receive in typical large-scale measurement studies.

Two years ago Nancy Bayley (5) addressed this Conference on the shape of the mental growth curve. She raised the question of what kinds of emotional climate are optimal at what ages, and what effects the attitudes of responsible adults such as parents and teachers have on intelligence. She suggested a few clues from some preliminary data on maternal attitudes. The second part of the Fels study (30) referred to earlier attempts to relate the changes in intelligence test scores in their longitudinal data to personality ratings of the children. Their hypotheses that the children gaining most in IQ ratings would show most favorable ratings on personality were confirmed. They state, "A study of the various modes of personality by which children attempt to gain satisfaction in their experiences appeared to be of value in predicting IQ change and in understanding the nature of accelerated or decelerated mental growth as related to personality factors. During the preschool years, emotional dependence on parents appears to be clearly associated with loss in IQ . . . During the elementary school years, a cluster of personality traits with the need for achievement as a common dimension appears to be closely associated with accelerated or decelerated mental growth patterns . . . during the preschool years the child who develops modes of behavior characterized by aggressiveness, self-initiative, and competitiveness is laying a basic groundwork for future acceleration in performance on mental tasks." (pp. 117-118) These authors conclude, then, that children who gained in IQ during the preschool years were those who were "venturing out of the maternal fold."

Goldfarb (13) cites Bowlby's (7) excellent summary of work on institutionalized children who, lacking maternal stimulation, affection and support, were retarded intellectually and "distinctly impaired in conceptual ability." The impairment in categoric behavior noted among institution children was considered to be more than a reflection of low intelligence. "There seemed to be a lack of differentiation and development of all aspects of personality. Most noteworthy was a generalized state of intellectual and emotional improvement and passivity. Along with the cognitive disability there were distinct emotional trends; chiefly, the absence of normal capacity for inhibition. The institution group showed extremely difficult behavior with symptoms of hyperactivity, restlessness, inability to concentrate and unmanageability. Further, although indiscriminatingly and insatiably demanding of affection, they had no genuine attachments. They were incapable of reciprocating tender feeling . . . there was an absence of normal anxiety following aggressive or cruel behavior (and) . . . specific impairment in social maturity."

This syndrome is in marked agreement with Lauretta Bender's (6) description of "Psychopathic behavior in childhood" which she characterizes by "an inability to love or feel guilty. There is no conscience. . . . There is an inability to conceptualize, particularly significant in regard to time. They have no concept of time so that they cannot recall past experience and cannot benefit from past experience or be motivated to future goals." This description calls vividly to mind Terman's (35) early description of what constitutes intelligent behavior, in which he said it is the ability to abstract out of past experience those essential features needed to meet new situations and the ability to adapt them in new situations.

REFERENCES

1. Anastasi, Anne. *Psychological testing.* New York: Macmillan, 1954.
2. Atkins, Ruth E. *The measurement of the intelligence of young children by an object-fitting test.* Minneapolis: Univer. of Minnesota Press, 1931.
3. Baldwin, A. L. *Behavior and development in childhood.* New York: Dryden Press, 1955.
4. Bayley, Nancy. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, 75, 165-196.
5. Bayley, Nancy. A new look at the curve of intelligence. *Proc. Invitational Conf. on Testing Problems.* Princeton, N. J.: ETS, 1956.
6. Bender, Lauretta. Psychopathic behavior disorders in children. In Lindner, R. M. and Seliger, R. V. (Eds.), *Handbook of correctional psychology.* New York: Philosophical Library, 1947.
7. Bowlby, J. Maternal care and mental health. *World Health Organization Monogr.*, 1951, No. 2.
8. Brodbeck, A. J., & Irwin, O. C. The speech behavior of infants without families. *Child Develpm.*, 1946, 17, 145-156.
9. Catalano, F. L., & McCarthy, Dorothea. Infant speech as a possible predictor of later intelligence. *J. Psychol.*, 1954, 38, 203-209.
10. De Hirsh, Katrina. Tests designed to discover potential reading difficulties at the six-year-old level. *Amer. J. Orthopsychiat.*, 1957, 27, 566-576.
11. Fisichelli, Regina M. A study of pre-linguistic speech development of institutionalized infants. Unpublished doctoral dissertation, Fordham Univer., 1950.
12. Gesell, A. *The mental growth of the preschool child.* New York: Macmillan, 1925.
13. Goldfarb, W. Emotional and intellectual consequences of psychologic deprivation in infancy; a revaluation. In Hoch, P. H. and Zubin, J. (Eds.), *Psychopathology of childhood.* New York: Grune, 1955.
14. Goodenough, Florence L., & Maurer, Katharine M. *Mental growth of children from two to fourteen years.* Minneapolis: Univer. of Minnesota Press, 1942.
15. Griffiths, Ruth. *The abilities of babies: a study in mental measurement.* New York: McGraw Hill, 1954.
16. Guilford, J. P., & Zimmerman, W. S. *The Guilford Zimmerman temperament survey, manual.* Beverly Hills, Calif.: Sheridan Supply Co., 1947.
17. Haeussermann, Else. Evaluating the developmental level of cerebral palsy preschool children. *J. genet. Psychol.*, 1952, 80, 3-23.
18. Harrison, L. M. Nature and development of concepts of time among young children. *Elementary Sch. J.*, 1934, 34, 507-14.
19. Hiskey, M. S. *Nebraska test of learning aptitude for young deaf children, manual.* Lincoln, Nebr.: Univer. of Nebraska, 1941.
20. Hofstaetter, P. R. The changing composition of "Intelligence:" a study in T-technique. *J. genet. Psychol.*, 1954, 85, 159-164.
21. Honzik, M. P. The constancy of mental test performance during the pre-school period. *J. genet. Psychol.*, 1938, 52, 285-302.
22. Irwin, O. C. Speech sound elements during the first year of life, a review of the literature. *J. Speech. Disord.*, 1943, 8, 109-121.
23. Klatskin, Ethelyn, H. Intelligence test performance at one year among infants raised with flexible methodology. *J. clin. Psychol.*, 1952, 8, 230-237.
24. Landreth, Catherine. *The psychology of early childhood.* New York: Knopf, 1958.
25. Leiter, R. B. *Manual for the 1948 revision of the Leiter international performance scale.* New York: Psychol. Service Center Press, 1948, Part II.

26. **McCarthy, Dorothea.** Language development in children. In Carmichael, L. (Ed.), *Manual of child psychology*. New York: Wiley, 1954.
27. **Martin, W. E.** (Chairman). Report of fact finding committee on the present status of training and research in child and adolescent development. *Newsletter Division on Developmental Psychology*. Amer. Psychol. Ass., Spring, 1957.
28. **Maurer, Katharine M.** *Intellectual status at maturity as a criterion for selecting items in preschool tests*. Minneapolis: Univer. of Minnesota Press, 1946.
29. **Sloan, W.** The Lincoln-Oseretsky motor development scale. *Genet. Psychol. Monograph*, 1955, 51, 183-252.
30. **Sontag, L. W., Baker, C. T., & Nelson, Virginia L.** Mental growth and personality development. *Monogr. Soc. Res. Child Develpm.*, 1958, 23, No. 2.
31. **Sontag, L. W., & Baker, C. T.** Mental growth and personality development, a longitudinal study. *Monogr. Soc. Res. Child Develpm.*, 1958, 23, No. 2.
32. **Templin, Mildred C.** *Certain language skills in children, their development and interrelationships*. Minneapolis: Univer. of Minnesota Press, 1957.
33. **Terman, L. M.** Intelligence and its measurement. *J. Educ. Psych.*, 1921, 12, 123-147 and 195-216.
34. **Thorndike, E. L.** *The measurement of intelligence*. New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1927.
35. **Thurstone, L. L., & Thurstone, Thelma G.** *Tests of primary mental abilities for ages 5 and 6*. Chicago: Science Research Associates, 1953.
36. **Wechsler, D.** *Wechsler intelligence scale for children*. New York: Psychol. Corp., 1949.
37. **Wyatt, Gertrude L.** *Speech and interpersonal relations*. [Personal Communication].

Prediction of Maladjustive Behavior

William C. Kvaraceus, Director Juvenile Delinquency Project, National Education Association*

Juvenile delinquency is a complex and contentious topic constantly fermenting on the American scene. Faced with the delinquency problem at the local level, a community can hope to achieve some measure of prevention and control only to the extent to which its efforts are characterized by the following:

1. *A positive community attitude.* The delinquent must be viewed as a child needing understanding and help rather than punishment and placement outside the community. It is generally true that the delinquent is a hostile child, but one who also faces an equally hostile community. We can't help the delinquent or his family doing business on a two-way street of hate and hostility.

2. *A knowledge of the delinquency phenomenon.* The community must plan on the basis of plausible and research-oriented theory of delinquency as a form of adjustment in our culture and subcultures; it must have some knowledge of the geography, psychology, and sociology of the delinquency act on the local scene; and it must come into the more intimate knowledge of the specific delinquent act through case-study approaches. Lacking knowledge at these three levels, the community may booby-trap itself into "impractical-practical" approaches (curfew, wood-shed, anti-parent legislation) which are irrelevant, if not harmful, to delinquency control and prevention. Common sense opinions cannot be trusted in this field. We must look for, and stick to, the data available from within the behavioral sciences—even in the face of the irrational (unscientific) lay critic, the frightening and shrill cry of the feature writer, or the crusading editorial commentator.

3. *Early identification of pre-delinquent and delinquent.* Delinquent behavior is not a 24-hour malady; it develops over a long period of time and usually with the generous assistance of two or three adults. The future delinquent presents many hints and rumblings of his coming explosions.

4. *Early referral for study and diagnosis.* Once the delinquent or pre-

---

*On leave from Boston University, 1958-59.

delinquent has been identified there must be available special service personnel (psychologist, social worker, psychiatrist, counselor) to help get at the meaning and predisposing factors of the maladapted behavior. True, these professional services cost money, but this is no dime-store problem.

5. *Treatment through coordination of community resources.* Using all child and family-serving agencies, the resources of the community must be brought to bear on the individual in a systematic, scientific, and individualized effort of rehabilitation based on case and community study.

This discussion focuses on the third aspect, on methodology and techniques aimed to help the community worker in the process of early identification of the pre-delinquent. Apart from the primary and direct attack on the delinquency problem via the general improvement in patterns of family living, in more effective school programs, in neighborhood value systems, and in leisure time offerings, etc., delinquency prevention programs will depend heavily on the ability to identify at an early date (perhaps as early as the first grade level) the youngster who is prone, vulnerable, exposed, or susceptible to the delinquent pattern of adjustment.

Contrary to the usual depressed predictive validity coefficients reported in the literature on forecasting success or failure in classroom achievement or on the job, I am happy to report that it is possible to predict with 100 per cent efficiency the future delinquents in our society.

Considering the complexity and the pressures of modern living within the enveloping web of social taboos, regulations, town bylaws, city ordinances, state and federal laws against the fact that there is still so much of Adam left in all of us, we can safely predict at least one good delinquency (usually more) for every man. (Witness, if in doubt, the mores of any out-of-town conventioning groups such as ours.) Everyone experiences several delinquencies, officially or unofficially, during the growth and maturation process. Even the road to sainthood often seems to have been paved with sins judging from the lives of many of those eventually beatified. But all this only raises the crucial question: "*Whom or what are we forecasting?*" I shall readdress myself to this query later. First, what are the instruments currently available on the market and how effective are they?

There are at least seven instruments or techniques which are available to the test user and which offer some claim, and sometimes some data to warrant mention—if not use—for early identification of the predelinquent. These include the following:

Personal Index of Problem Behavior (1)

Minnesota Multiphasic Personality Inventory (2)
Porteus Maze Test (3)
Washburne Social-Adjustment Inventory (4)
Clueck Prediction Tables (5)
Behavior Cards: A Test-Interview for Delinquent Children (6)
KD Proneness Scale and Check List (7)

None of these items is infallible, nor has any one of these methods demonstrated sufficient forecasting efficiency or power to be used in a routine or perfunctory fashion. The best that can be said for some of them is that they are promising and that they merit perhaps another master's, if not doctorate, thesis by way of further or partial validation.

Rather than review the content of their manuals which are available to any discriminating test user, I shall focus on the factors which tend to raise or to lower the reliability and the validity of such instruments. First, I shall discuss the basic premises or assumptions on which prediction methodology in the delinquency field is generally based, although not always acknowledged. Second, I will review some special construction and validation problems that must be solved if delinquency prediction is to become a useful and practical reality, rather than a hopeful research fantasy.

Basic Premises

Continuity of behavior. In child study and rehabilitation there is ever present the backward look to earlier life experiences of the subject in an effort to unlock the meaning of behavior or misbehavior. Stated in popular, if not poetic, tongue, past is prologue to the future, the child is conceived as father to the man, and concern is expressed with how the twigs have been bent, if not pruned. Prediction assumes a continuity in behavior or misbehavior, if you will, linked in a cause-effect sequence that is visible or discernible to an observer. However, the sequitur of cause and effect may not be visible to the naked or untutored eye. Most observers today view behavior casually through the distortion of their own bifocals, thus reflecting the bias of their own theoretical frame of reference. The forecaster of delinquency might thus overemphasize or underemphasize data obtained through somatotyping, psycho-genic data, psychoanalytic data, or sociological information that might pertain to the ultimate effect as seen in delinquent behavioral adjustment. It is my conviction that much of the continuity of cause and effect in delinquency is to be found in the cultural and subcultural stream, rather than embedded among factors under the skin, although the careful listener will note the implied false dichotomy. Any forecaster today cannot afford to overlook the cultural aspects of the delinquency phenomenon.

If some delinquency is spontaneous and accidental, this premise is weakened, or poorly maintained; if some youngsters prepare for their delinquencies quietly and pleasantly, as appears to be the case in a growing number of cases, the observer, whatever his theory, may be hard put to spot the future offender.

Factor modality. The forecaster assumes that there is a factor modality among enough variables that are commonly and peculiarly associated with delinquent behavior. These factor modalities represent significant differences that pile up between those who become delinquent and those who do not resort to this adjustive mechanism. On the other hand the singular and unique nature of each offender's syndrome tends to deny and to demolish any build up of a useful common modality on which to base a forecast of malbehavior. At the same time, isolation of a number of commonly observed variables in the backgrounds of delinquents as contrasted with nondelinquent counterparts involves an isolation and atomizing of elements that sacrifices dynamic aspects in causal relationship, hence reducing forecasting efficiency.

Unreliability of stimulus variables in class and subculture. Whatever factor, stimulus, or variable is selected for use in a prediction scheme, it is likely to fall victim to differential interpretation according to the respondent's value system reflecting the ways of thinking, behaving, adjusting in the subculture or class with which he is identified. Hence, love and affection may be conceived as tender and be identified with a lullabye in the upper class; love perceived by a lower class adolescent may be viewed as fierce and violent and be identified with a family fight, thus proving the worth and importance of the young member in the family arena. A good example can be found in the use of the affection item in the Glueck social factors' table with a Maltese father, who culturally never displays open affection for his young, although the table expectancy is that he should do so. School achievement may be slurred in the lower class home and praised in the upper class family, duplicity and cunning may be extolled in lower class living and deplored in upper class membership.

If delinquency is an essential and more typical aspect of life among the lower classes, these differential responses can be exploited in the forecasting game. However, as more delinquents tend to be drawn from the upper levels of community structure in the future, these stimuli and variables which invite differential responses among youth may only succeed in sorting them into the classes and/or subcultures from which they come, or with which they most easily identify. The value of such class identification would thus be lowered for prediction purposes.

Contingency in predicting behavior. Predicting future adjustments or maladjustments will always be made on a contingency basis. If the subject's situation improves, the prediction of delinquency will be weakened; if the situation deteriorates, the forecaster is more likely to be right this time. Hence the prediction made in time must be viewed as relative to subsequent conditions. Failure to predict accurately may often reflect validation of the methodology albeit it depresses the validity coefficient.

Short-term vs. long-term prediction. Some forecasters, particularly the Gluecks, have assayed long-term prediction working with the six-year-old or from the first grade level. Just as it is hazardous to plan a picnic or a skiing trip on the basis of long term weather forecasting, one must be prepared for disappointments as well as surprises. Obviously predicting at the junior and senior high school levels, closer to the point of delinquency precipitation, should yield a higher level of validity coefficient than in long range forecasting. A pertinent question arising is that of adequate time-allowance in the validation of any prediction scheme to insure adequate measure of delinquency fall-out. Current British studies reported by Mannheim suggest that 18 months may be sufficient to check the prediction power of some measures with older youth who have been institutionalized. Beyond this we are in the dark about what constitutes minimal time duration in an effective validation design.

Such are the major premises on which prediction techniques are set. Some of these assumptions may withstand close inspection while others cannot be accepted axiomatically. Hence, we do not have a firm bedrock on which any prediction scheme can be automatically and easily constructed.

Special Methodological Problems in Construction and Validation

Assuming that a workable base can be squared off, the forecaster must still face the following special problems in the test construction and validation process.

Validation design. There is no substitute for the before-and-after research design in validation studies of prediction tools. This means that the prediction technique must be applied to a sample of youngsters and forecasts made. A reasonable period of time for behavior and misbehavior to take place must be allowed during which adjustment criterion data must be gathered according to some acceptable definition of malbehavior. Finally, the relationship between the forecast and the behavioral adjustment must be established and expressed in terms of prediction efficiency.

Many of the techniques now on the market depend too heavily, even exclusively, on construct validity or concurrent validity. Many validation studies merely telescope the before-and-after design by using direct comparisons between available criterion groups and still others attempt to validate forecasting effectiveness via retrospective analysis. Throughout many of these studies there is some confusion between what constitutes probability and what denotes predictive efficiency in a statistical design.

Before-and-after studies are expensive and difficult to manage. Loss of cases due to mobility alone is a serious stumbling block to the research in such long-term experiments. But there is neither a haven nor excuse in these difficulties for poorly executed research.

The criterion: Who or what is being predicted? To return to the question raised earlier: "Whom or what are we predicting?" What kind of criterion data are to be collected on each individual in the post-forecasting situation? Who and what is a delinquent? The term, juvenile delinquent, is an omnibus concept that can include the large bulk of our youth population since everyone can, and does, easily fall by the wayside at one time or another in our more complex Garden of Eden.

We must first observe that there is no dichotomy between delinquents and non-delinquents (except in terms of the court tag, but even here the dichotomy breaks down as one studies the informal and formal dispositions of cases). The implication is that the statistical design in prediction of malbehavior and delinquent behavior is not amenable to the expediency of a biserial r. Misbehavior exists on a continuum. What the researcher lacks is a graduated measure of the delinquency phenomenon on a malbehavior scale. Until such a measure, based on some system of habituation and seriousness of offense, is worked out the unreliability of the criterion measure itself will seriously reduce the validity coefficient assuming that a high degree of relationship exists between forecasting technique and adjustment.

All existing forecasting devices have attempted to predict to the galaxy of any and all kinds of delinquency without due regard for any diagnostic differentiation according to modalities (types) of delinquents. This is perhaps their greatest defect. Separate validation checks and prediction tables need to be evolved for the following types or modalities: the *neurotic delinquent*, heavy with anxiety and guilt; the *socialized delinquent* who has failed, for good reason, to internalize the value system of dominant society and whose super-ego is already delinquency identified; the overt-aggressive *unsocialized delinquent* whose behavior represents a strong defense, even offense, against authority figures conceived as hostile, threatening, and predatory. To these major modalities might be added others, *viz.* group intoxicated type, traumatized delinquent, constitutional type, and perhaps others.

Refinement in validation experiment will await refinement in differential diagnosis in the process of gathering criterion data against which to test the forecasting instrument, always working within the rubric of each modality. I would hazard the hypothesis that we can succeed in predicting certain modalities of delinquents with greater effectiveness than others. For example, it may be relatively easy to identify the future socialized and unsocialized delinquents and relatively difficult to predict in the neurotic category of delinquents. At the same time it may prove an impossible task to predict the traumatized offender because of the accidental nature of this phenomenon.

One additional note bears mentioning. The popular practice of resorting to the use of such convenient and undifferentiated criterion groups as court cases or, worse, institutionalized delinquents on whom the community has given up or who have been removed from the community for special reasons, should be erased from validation studies which aim to set up prediction tables. The special and hardy breed of screened delinquents obtained through such sampling do not lend themselves to fair test or experimentation. In a sense the cards are stacked in our favor. Without the use of any elaborate device most youth workers in any community can foretell what youngsters are most likely to be banished to the training institutions.

Need for local validation. The delinquency problem varies from one community to another and in the large urban centers it will vary from neighborhood to neighborhood. Each community or neighborhood will show significant variations in incidence, type, and time of misbehavior, reflecting unique elements in the population and in the culture and/or subcultures. Any prediction tool that has been demonstrated as useful in a large urban center with a mixed or heterogeneous population may prove to be of little or no value in the more homogeneous and monolithic culture of suburbia. Any promising instrument now available on the market needs to undergo local validation rechecks. This will call for considerable research interest, effort, and skill on the part of test users at the local level.

The need and necessity for local validation of prediction techniques is also pointed up by the problems presented in the class structuring of American society. Sociological studies infer that the average New England community, for example, might need at least three separate editions of a prediction scale or table: one for the lower class, one for the middle class, and one for the upper class. Rather than hoist six editions to suit Warner's family status studies, I have compressed the categories perhaps too conveniently. On review I might prefer four editions. As one studies the items on some of the scales and check lists, the obvious irrelevancy of many of the stimuli for children of varying family status in our society is such as to render them useless and meaningless.

Observation vs. test situation. In developing a methodology of prediction we will need to favor the use of observation techniques such as check lists, graphic rating scales, anecdotal records as against the use of test items or self-inventory questionnaires which place a much too heavy burden on the reading ability, trustworthiness, and seriousness of purpose on the part of the respondent. The combination of low reading capacity, irrelevancy of response, and cultural duplicity of many pre-delinquents often tends to lower the reliability of the best of these instruments.

Furthermore the technique that is evolved must be easily administered to large classroom groups. To build prediction tables assuming Rorschach testors, Psychiatric Interviewers, and trained Social Workers will not result in any usable or practical detection methodology as we contemplate ten million youngsters in the high schools of the nation. What we need, for example, is a handy method which trained teachers can employ as they come in close and continued contact with their students.

Whatever instrument is devised, it must face the practical test of serving as an improvement over what might be accomplished even now through the careful reading of a case-study folder or a cumulative record file. The professional worker, clinically trained in child development and adolescent psychology, can generally smell out a future delinquent through a careful perusal of a child's case record. Prediction methodology must provide a shorthand method that is at least as effective as this longhand approach.

Summary Statement

How much hope can be extended to the community workers who are concerned with prevention and control of malbehavior through early identification of the potential nonconformer? If we can strengthen the base on which prediction methodology must be built through a careful re-analysis of all our major premises, and if we can carry off our construction and validation processes with the refinements which have been indicated, it is likely that we can predict delinquency as well as tests of academic aptitude predict academic achievement—which, as you all know, is not very well. Even then, or especially then, we shall need to recognize that the prediction tool and the data gathered thereby have in no way relieved us of the urgency and the necessity of careful and deliberate judgment in drawing conclusions, using all other available information, concerning the child's exposure or proneness to malbehavior and delinquency.

REFERENCES

1. GRAHAM C. LOOFBOUROW AND NOEL KEYS, *The Personal Index*. Minneapolis: Educational Test Bureau, 1933. WINIFRED C. RIGGS AND ARNOLD E. JOYAL, "A Validation of the Loofbourow-Keys Personal Index of Problem Behavior in Junior High Schools," *Journal of Education Psychology*, XIX (March, 1938), 194-201.
2. STARKE R. HATHAWAY AND C. CHARNLEY MCKINLEY, *The Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation, 1943. STARKE R. HATHAWAY AND ELIO D. MONACHESI, *Analyzing and Predicting Juvenile Delinquency with MMPI*. Minneapolis: University of Minnesota Press, 1953.
3. S. D. PORTEUS, *Qualitative Performance in the Maze Test*. New York: Psychological Corporation, 1942.
4. J. W. WASHBURNE, *Washburne Social-Adjustment Inventory*. Yonkers-on-Hudson, New York: World Book Co., 1938. J. W. WASHBURNE "An Experiment on Character Measurement," *Journal of Juvenile Delinquency*, XIII (Jan. 1929), 1-8.
5. SHELDON GLUECK AND ELEANOR GLUECK, *Unraveling Juvenile Delinquency*. New York: Commonwealth Fund, 1950.
6. RALPH M. STODGILL, Behavior Cards: *A Test-Interview for Delinquent Children*. New York: Psychological Corporation, 1949.
7. W. C. KVARACEUS, *KD Proneness Scale and Check List*. Yonkers-on-Hudson, New York: World Book Co., 1953.

Session II

Remarks of the Chairman

In this second session of the Conference, we shall hear reports that differ appreciably in content and approach from those presented in the first session. With diversity as our keynote, this is to be expected.

Those of us who have been concerned with test development and test construction are aware of some of the very notable and significant contributions to test theory that have come from the lady who will speak first. She is Dr. Jane Loevinger, of the Jewish Hospital of St. Louis, who will discuss "A Theory of Test Response."

Knowing full well that some of the points Dr. Loevinger raises will stimulate a desire for discussion—and knowing equally well that discussion time is limited—we have resorted to the strategy of asking one person to voice his reaction to her thesis. He is Dr. David V. Tiedeman, of the School of Education at Harvard University. Speaking as an individual, and not in any sense serving as surrogate reacting for all of us, Dr. Tiedeman will discuss Dr. Loevinger's theory immediately following her presentation.

Then, still bound by time limitations, we shall proceed to the next topic on our program. Earlier this morning we heard a discussion on the prediction of maladjustive behavior among children. In this session we shall hear a report on research in another important area of prediction, that of the "Measurement and Prediction of Teacher Effectiveness." Dr. David G. Ryans, of the University of Texas, is especially well qualified to bring us this report. He was Director of the National Teacher Examinations for a period of seven years, is currently Director of the Teacher Characteristics Study, and will draw extensively on the findings in that investigation in his presentation today.

Following Dr. Ryans' report, we shall have a brief comment from a man who is concerned, day in and day out, with precisely this matter of selection of teachers and the problem of how to identify good teachers. We are fortunate in having Dr. Harry B. Gilbert, of the Board of Examiners of the New York City Board of Education, here today to discuss some of the implications of Dr. Ryans' research findings.

37

A Theory of Test Response*

Jane Loevinger, Research Associate, Jewish Hospital of St. Louis

In recent months, working in collaboration with Professor Abel Ossorio and Mrs. Kitty LaPerriere, I found taking shape a conceptualization of a trait which I currently believe to be the major source of variance in structured personality tests, regardless of their intent. Manifestations of the trait have been called facade, test-taking defensiveness, response set, "social desirability," acquiescence, and so on. The term *response bias* can serve as generic for these phenomena, though Jackson and Messick (12) prefer to emphasize that they are components of personal style. The fact that response bias is a manifestation of the trait by no means implies that the trait is inconsequential outside of the testing situation. On the contrary, its importance in the test situation reflects its importance in many other aspects of life.

The trait may be defined metaphorically as the ability to assume distance from oneself, or more exactly, as capacity to conceptualize oneself. It is one cognitive aspect of ego development. That it should greatly influence the kind of self report which most personality tests call for is obvious; not quite so obvious is that capacity to conceptualize oneself varies as a function of age, of education, and of one's station in life. Note that the trait does not refer so much to the content of one's self-concept as to one's ability to form a self-concept. At least three points are needed to bring the dimension into focus. At the lowest point there is no capacity to conceptualize oneself; at the midpoint there is a stereotyped, usually conventional and socially acceptable self-conception; and at the highest point a differentiated and more or less realistic self-concept.

Let us look first at how this trait normally develops with age. We are all familiar with the baby's wonder as he discovers his own body. But at stake here is the conception of one's psychological rather than one's physical person. The moment of discovery may be perhaps the time the child first says "Bad" to himself as he does or refrains from doing something his parents have proscribed. At that moment the child has conceptualized himself as having impulses which are sometimes bad and are

---

*Preparation of this paper was supported by a research grant, M-1213, from the National Institute of Mental Health, Public Health Service.

not necessarily acted on. That he has achieved a rudimentary idea of good and bad is no more important than that he has a rudimentary idea of impulse and control. Extension and elaboration of the pair of constructs, impulse and control, are major tasks all through childhood. At this point the argument is reminiscent of Kelly's (14) psychology of personal constructs: to have the construct of impulse is to have its opposite, control, and thus to achieve a degree of choice. What Kelly does not seem to recognize clearly is that in just this instance, not having the construct of impulse does not eliminate impulses from one's repertory but rather leaves one completely at their mercy.

By early adolescence the ability to conceptualize one's impulses and the concomitant degree of control is fairly well established. But the typical adolescent is in many respects an "authoritarian personality" (1). He is prone to think in stereotypes, to be punitive, disciplinarian, conventional, anti-psychological and intolerant of those who are different (8, 18). In terms of the aspect of ego development here being described, he has achieved distance from his impulses but not from his ego. He has some ability to think about himself as a psychological person, but his self-characterization tends to follow a conventional, socially approved stereotype. The strongly derogatory self-portrait which is also common in adolescence is equally stereotyped. Everyday observation leads to the suspicion that the derogatory and flattering stereotypes may alternate in some children in short time span.

During the college years there is in favorable instances a change from the typically authoritarian to an intellectually sophisticated point of view. There is usually, or at least often, a marked increase in capacity to view oneself with some detachment, to see oneself as having a style of life, to report feelings without taking refuge in conventional stereotypes. The changes which take place during the college years in personality in general and test behavior in particular have been documented in the Vassar study of Sanford, Freedman, and Webster (21). That response stereotypy on tests is somewhat characteristic of college freshmen but not of advanced undergraduates has been noted by Christie, Havel, and Seidenberg (4).

Thus the capacity to attain distance from oneself grows with age, from infancy, where there is no distance from impulse, through adolescence, where there is distance from impulse but not ego, to the college years, where there is distance from ego as well as impulse. Were the sole purpose of this discussion to present a picture of the normal course of ego development, this would be a pale and one-dimensional version of Erik Erikson's (8) vivid and dynamic portrait, "Growth and Crises of the

39

Healthy Personality," prepared for the Midcentury White House Conference. My purpose is not to describe personality development but to make a contribution to personality measurement. Erikson's brilliant paper does not, by itself, lead to measurement. The elision of some of the stages Erikson describes results from using available psychometric research as a sieve for his intuitive observations.

The usefulness of the concept of ego development as a psychometric dimension depends on whether one can convincingly describe some individuals in terms of levels of ego development not characteristic of their age. When we speak of an adult as having a mental age of 2 or 8 or 10, we of course do not mean that his behavior is identical with that of the average child of the given age. Rather there is an abstract characteristic of his behavior which can be so measured. Similarly, some children and a few adults have as little control and as little capacity to conceptualize their impulses as infants or small children. Redl and Wineman (17) have depicted preadolescent children of this type in a book which contributes much to our understanding of ego development. Many, perhaps most, adults have a self-conception hardly less stereotyped than that characteristic of adolescence. If we take a slice of the population of constant age, we will find ego development as measured on this hypothetical scale correlated with intelligence, with educational level, and with some measures of social class. Since intelligence, social class, and educational level are themselves intercorrelated, this represents a single additional datum in support of the conceptualization. But while it is a single argument, it is supported by a large amount of research. Several recent summaries of research with the California F scale, which is as much a measure of ego level as of anything, have confirmed these relationships (3,5,23).

There have been now presented three lines of argument in support of the construct of ego development. It is a constantly increasing function of age, at least through the early adult years. It tends to increase constantly as a function of intelligence, educational level, and social status. And it can be conceptualized as increase in a single function, to wit, capacity to assume distance from oneself. There is a fourth argument: ego development tends to increase constantly with psychotherapy.

All forms of psychotherapy push the patient upward on this dimension. In the case of delinquent or disorganized persons who remain at the low end of the scale beyond the appropriate age, increase in conventionality and control is an aim. In the case of conventional people, increase in sophistication, in capacity to conceptualize themselves, is not necessarily an aim but is the means of therapy. Rogers (20), in describing the therapeutic process, has postulated essentially the same dimension. An apparent paradox is that Rogers describes the later stages of therapy in terms of *decreased* distance from one's feelings. This paradox is explained in psychoanalytic writings in terms of a temporary splitting of the ego (7). To achieve distance from oneself is the condition for achieving immediate grasp of feelings as feelings. The patient must talk about himself and his feelings in the therapeutic transaction, which surely implies a capacity to attain distance from them.

Let us look at characteristic manifestations of different levels of ego development in personality tests. The most striking fact, of course, is the tendency of most people to answer in terms of response stereotypes, most notably, a defensive or favorable self-portrayal. The tendency to describe oneself favorably has been shown to increase between the ages of 8 and 13 (10) and to decrease during the college years (21). This non-monotonic relation between "socially desirable" self-portrayal and age corresponds exactly to the non-monotonic relation between conventionality and ego development. That is, conventionality tends to increase as we go from lowest to middle level, and to decrease between middle and highest level. The non-monotonic relation between the most obvious phenotypic test manifestation and the genotypic trait is surely a major obstacle to personality measurement.

In measuring maladjustment, neuroticism, and the like, one must set up a set of responses as "normal." Psychologists in recent years have tended to use a statistical definition of normality. So doing, however, does not alter the normal key very much from what would have been chosen a priori by psychologists of a more naive era, for the socially acceptable responses are just the ones chosen most frequently by the numerous middle group. Middle-class children and adults tend to appear a little better adjusted on personality tests than their lower-class contemporaries (2). On the other hand, Vassar seniors tend to test more maladjusted than Vassar freshmen (21). While personality changes undoubtedly take place during the college years, they are probably predominantly in the direction of greater ego development and intellectual maturity. Its seems unlikely that basically the seniors are a lot more maladjusted than they had been as freshmen. Rather, they are more self-critical, less conventional and stereotyped in their thinking. They are capable of admitting to consciousness and to their test responses problems which had been there all along but were concealed beneath a facade of normality. Just this sort of phenomenon has made measurement of adjustment enormously difficult. For psychotherapy itself tends to move the patient up the scale of ego development. And, other things equal, greater

41

ego maturity at the upper extreme leads to a decrease in stereotyped favorable self-portrayal. No doubt there are many exceptions to the latter generalization. Some sick people give a stereotyped unfavorable self-portrayal, and therapy could be expected to brighten their self-portrait. Moreover, symptoms which actually disappear could be expected to be so reported. All in all, however, evidence indicates that response to structured personality tests is more clearly related to ego development than to adjustment.

What the test behavior is of those lowest on ego development is not known in detail. These people include small children, as well as older children and adults in whom impulsivity is unduly predominant. Anyone who has tried to obtain tests from individuals of low social status, where the lowest level of ego development is overrepresented, has discovered that refusals to cooperate and sabotage of various sorts are more frequent than at higher social levels. The suspicion of and resistance to such small authority as a research psychologist represents is itself a fact worth recording, and strikingly similar to the negativism of pre-school children in testing situations.

These individuals do become accessible to psychological observation through more or less involuntary referral to guidance clinics, alcoholic treatment centers, and so on. Their inability to put their troubles, however overwhelming, into words is only partly a matter of opposition to the authority of the clinic. Skillful, sympathetic clinicians report that it appears to represent a genuine inability to conceptualize themselves. Diffuse physical complaints seem to represent a kind of "body English," i.e., their physical complaints may represent in part psychological malaise for which they have no concepts.

Resistance to authority, impulsivity, and lack of ability for self-conceptualization: surely this is a coherent syndrome, and one different from the identification with authority which characterizes the midpoint of our variable. Documentation of this syndrome can be found in the description of the lowest social class by Hollingshead and Redlich (11, Ch. 4), though I do not maintain that all individuals in the lowest class are at the lowest level of ego development.

The ideas presented here are meant to apply chiefly to objective personality tests. There is, however, one problem of long standing in projective testing, especially the Thematic Apperception Test (TAT), which has some relation to these concepts. The puzzle is, when do aggressive responses on the TAT indicate aggression in overt behavior, and alternatively, when do aggressive fantasies substitute for aggressive behavior? Probably no one has a complete and clear-cut answer to this question. There are indications that in youths from low social classes there is a positive relation between overt and fantasy aggression; in higher social classes, a negative relation. The psychoanalytic concept of "primary process" helps to bridge the gap between this finding and the concept of ego development. Predominance of primary process is a translation of the impulsivity which characterizes the lowest level of ego development. For individuals at this level words and fantasies serve to trigger the kinds of behavior they symbolize. Since there is minimal control of impulse expression in behavior, the impulses expressed in fantasy are the same as those expressed in behavior. In the two higher levels of ego development, on the contrary, secondary process is well established, which is to say, words and fantasies serve to delay, control, and substitute for expression of impulses in behavior. Therefore, it is not surprising to find a slight negative relation between fantasy aggression and overt aggression in middle and upper class groups.

Lyle and Gilchrist (15) compared TAT protocols of delinquent boys with those of a matched control group; there was no difference in the number of aggressive or anti-social themes expressed, but the non-delinquents used various devices to indicate greater distance from anti-social impulses, such as denial of the reality of the situation, inhibition of the impulse by guilt, and rationalization of the anti-social act. Note that Lyle and Gilchrist use the same metaphor, distance, to indicate the means by which control over impulses is maintained, and that they find representation of the control devices in the TAT protocol. Purcell (16) found similar results studying psychiatric referrals in an Army training camp. He divided his cases into three groups according to case history evidence of anti-social conduct. Best differentiation of the groups was in terms of fantasy themes of internal punishment and ratings of the aggressive fantasies as to "remoteness," which referred to time, place, degree of reality, and so on. The anti-social group showed few themes of internal punishment and little remoteness from their aggressive fantasies. While Purcell interpreted the absence of themes of internal punishment in the anti-social group in superego terms, note that it is also evidence of lack of ability to conceptualize inner life.

Dr. Kenneth Isaacs has just sent me a MS. in which he develops a construct very similar to what I call ego development. He calls it "relatability," stressing the capacity to perceive other people and capacity for differentiated interpersonal relations. His studies also show that level of relatability can be judged from TAT protocols, but specific criteria are not listed.

In sketching ego development as a major dimension of personality, I am using Binet's work as my model. His breakthrough in the field of ability measurement, which has not been matched in the following 50 years, succeeded, I believe, because he found a *process* which corresponded to an intuitively perceived *dimension*. There have been attempts to define personality traits after Binet's model. But if we imitate him too closely, we end up measuring almost the same trait that he did rather than a personality trait. In fact, correlation of a personality test with age or intelligence is often interpreted as that much evidence for invalidity. Yet it is absurd to assert that personality does not change with age, and both gratuitous and contrary to everyday observation to assume that personality trends will be uncorrelated with IQ or social status. We cannot lift ourselves out of this problem by our correlational bootstraps; we need to mind the psychological content of our measurements.

The dimension of ego development I have sketched is a kind of common denominator in Erikson's description of the normal process of ego development; Rogers' description of the process of therapy; Sullivan, Grant, and Grant's (22) description of the growth of capacity for interpersonal relations; and results with objective personality tests pertaining to authoritarianism and response stereotypy. The three papers describing process all refer to seven stages, whether by coincidence or not. From the psychometric point of view, each involves a forbidding array of details. On the other hand, the psychometric approach has been to assume that everyone can be classified as having more or less of some one thing, like dominance or adjustment, or lies somewhere between a pair of poles, like authoritarian-democratic. The idea that if you can name it, you can measure it, dies hard. The journals are full of studies using little *ad hoc* tests of traits that struck that research worker's fancy.

I have, then, followed Binet in using process as touchstone of dimension; but I have tried to avoid the circumstantial details of particular processes, as well as the nominalistic fallacy which still vitiates many psychometric approaches to personality.

The authors of *The Authoritarian Personality* (1) considered and rejected the idea that the authoritarian was an immature version of the liberal person. However, the meaning of authoritarianism shifted in the course of their research. At its inception they were concerned with a harsh and pathological extreme of anti-Semitism and fascism, which they found only in a few individuals in their San Quentin sample, who fit the description of the lowest level of ego development. The core of the trait which emerged from their studies was very much like what I have described as the middle stage of ego development, much less vicious than what they looked for at first. The reason the California group was diverted away from the political aspects of authoritarianism and in the direction of ego development is that the latter aspects are far more pervasive in personality and are just the aspects of personality most accessible to measurement and interview.

The California disclaimer that authoritarianism and liberalism are stages in a developmental process has not stood up. Evidence that adolescence is typically a more authoritarian period than later maturity has come from many independent sources, including clinical observation (8), opinion polls (18), and studies with the F scale. The thinking of the California group evolved from that of looking for a few wicked authoritarians to recognizing the authoritarian tendencies in large groups of ordinary people. They never quite admitted that the conventional authoritarian represents the norm in our society. To see authoritarian tendencies in a developmental framework, as I have tried to do today, is to carry the evolution of the concept one painful step further: The struggle against authoritarian tendencies is one which each of us must make within himself, and it is a battle never wholly won.

That the fight against authoritarianism takes place in each of us was the theme of Erich Fromm's (9) 1941 book, *Escape from Freedom.* But Fromm, seeing the similarity in the child's spontaneity and the spontaneity which can be recaptured by a truly mature adult, wrote as if people knew their real selves and then deliberately surrendered that knowledge to slip into a conformist or authoritarian stereotype. The dialectics of growth seem more accurately represented by the sequence: impulsivity, rigid control enforced by intellectual stereotypes, and flexible controls enforced by genuine insight. Riesman (19), though much influenced by Fromm, has drawn a picture essentially the same as that sketched here. His term for the lowest level of ego development is anomic; for the middle level, conformist or, most often, adjusted; for the highest level, autonomous. Riesman has enriched our understanding of different patterns of conforming by his description of the tradition-directed, the inner-directed, and the other-directed man. These types of conformity characterize the middle level of ego development in different societies and in different groups within a given society. So far no one has traced the differential manifestations on tests of the inner-directed and the other-directed man, though there has been at least one attempt. But Allen Edwards' (6) finding, that the number of people claiming that an item describes them is a high rectilinear function of the independently judged "social desirability" of the item is remarkable evidence for our other-directedness.

Proponents of factor analysis, cluster analysis, and multidimensional scaling set up artificial problems with boxes or random numbers and demonstrate that their preferred method will indeed capture the dimensions built into the problem. I have begun instead with a real trait of central importance in test behavior and would now emphasize that factor analysis or cluster analysis or multidimensional scaling could not possibly reconstruct such a trait. Impulsivity is a distinguishing mark of the lowest level of ego development, but the flexible controls of mature life are phenotypically closer to the impulsive stage than are the rigid controls of the intermediate stage. The non-monotonic relation between conventionality and ego development has already been noted. The complex of ego development leaves many traces, and with respect to each of them there are individual differences. Factor analysts make much of getting from phenotypic variables to genotypic traits. But only such genotypic variables as are linearly, or at least monotonically, related to phenotypic ones will be revealed by factor analysis. By themselves, statistical techniques can yield only partial insights. I trust, however, that no one will carry away the message that I don't think it worthwhile to master or use difficult statistics. The psychological research worker who does not understand statistical principles is as handicapped as the psychometrician who does not permit himself to develop a feeling for the traits he studies. Factor analysis is an important technique in the hands of a responsible psychologist with insight into the psychological content of his variables.

Suppose you answer that you prefer to stick with whatever factor analysis reveals, that you find nothing compelling about the construct I have sketched. This raises an interesting and profound question, one which will be answered neither in short time nor by the self-elected. Since personality is complicated enough to encourage many alternative constructions, what are the criteria for the validity of alternative ways of construing it? Ego development as here sketched provides a framework within which one can view such major researches as *The Authoritarian Personality* and subsequent related studies (1,3,23); Redl and Wineman's (17) *The Aggressive Child;* Riesman's (19) studies of American character; Erikson's (8) work on growth and crises of the normal personality; studies of personality development in the college years by Sanford and others (21); Edwards' (6) work on the social desirability variable in personality tests; work on the relation between content and style or response bias (12); Rogers' (20) study of the process of therapy; and Kelly's (14) psychology of personal constructs. Dozens of smaller or less familiar studies contribute also to the overall picture. The line of research which originally gave rise to these speculations, work which I have been doing with Blanche Sweet, Abel Ossorio, Kitty LaPerriere, and others on patterns of child rearing, I have not even mentioned. Work now in progress is testing the hypothesis that different patterns of child rearing characterize different levels of ego development; here is another far-reaching application.

If memory serves, factor analysis originally aimed to give an economical account of much data with few concepts. I am claiming that the single construct I have proposed accounts for much data. By contrast, application of factor analysis to personality tests has too often taken small amounts of data and developed a confusingly large number of constructs. Has factor analysis of personality tests given rise to any powerful constructs, any constructs of sufficient utility, for example, that clinicians have made use of them?

A problem of concern to the Educational Testing Service has been measuring the behavioral outcomes of higher education. I would like finally to show how this problem is related to the discussion. In regard to education at the nursery school and kindergarten level, no doubt specific behaviors can be used to measure the success of the educational endeavor. The child is taught to lay his coat on the floor, slip his arms into it and flip it on by raising his arms. He must learn to conform to bells, commands, and classroom routines. The aim of university education is emphatically not to inculcate such stereotyped behavior patterns, but to free the graduate from conformity to cultural and behavioral stereotypes. I do not have any pat suggestions as to how to measure the outcome of higher education, but it seems safe to say that the search for specific behavioral outcomes is doomed to failure. It represents, moreover, a spurious and misguided objectivity. William James made the point in his essay on Harvard: "The day when Harvard shall stamp a single hard and fast type of character upon her children, will be that of her downfall. Our undisciplinables are our proudest product" (13, p. 355).

Summary

A cognitive aspect of ego development, ability to conceptualize oneself, is postulated as accounting for a major portion of the variance in structured personality tests. At least three points are needed to define the dimension. At the lowest point there is no capacity to conceptualize oneself as a psychological person; at the midpoint, a stereotyped self-conception; at the highest point, a differentiated, realistic self-

conception. More or less synonymously, at the lowest point there is no distance from impulses; at the midpoint, distance from impulse but not from ego; at the highest point, ability to assume distance from ego as well as from impulse. This trait increases constantly with age; for constant age tends to increase with intelligence, education, and social status; and tends to increase with psychotherapy. However, ego development has no conspicuous constantly increasing manifestations. Its most conspicuous manifestation in personality tests, tendency to answer in a stereotyped, usually a socially approved style, is not monotonically related to the trait, tending to decrease in the upper range and probably tending to increase in the lower range. A further difficulty in measuring favorable outcome of higher education, and incidentally, favorable outcome of psychotherapy, is that the highest level of ego development is characterized precisely by the absence of stereotyped, objectively specifiable behaviors and attitudes. I have followed Binet in using process as touchstone for dimension, but not imitated him too closely for fear of returning exactly to general ability.

Many methodologists, until recently including me, believe that our job is to perfect a method for discovering traits, and the right method will lead us straightaway to a complete catalogue of important traits. Purely for its shock value I wish to record a contrary hypothesis, that every major human trait will be discovered and established by a unique method. Whether that hypothesis is true or not, methodological sophistication in the absence of psychological acumen will lead only to fragmentary dimensions and insights.

REFERENCES

1. Adorno T. W., Frenkel-Brunswik, Else, Levinson, D. J., and Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. Auld, F. Influence of social class on personality test responses. *Psychol. Bull.*, 1952, 49, 318-332.
3. Christie, R., and Cook, Peggy. A guide to published literature relating to the authoritarian personality through 1956. *J. Psychol.*, 1958, 45, 171-199.
4. Christie, R., Havel, Joan, and Seidenberg, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, 56, 143-159.
5. Cohn, T. S. The relation of the F-scale to intelligence. *J. soc. Psychol.*, 1957, 46, 207-217.
6. Edwards, A. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.
7. Ekstein, R. Psychoanalytic techniques. In D. Brower and L. E. Abt (Eds.), *Progress in clinical psychology*. New York: Grune & Stratton, 1956. Pp. 79-87.
8. Erikson, E. H. Growth and crises of the healthy personality. In M. J. E. Senn (Ed.), *Symposium on the healthy personality*. New York: Josiah Macy, Jr. Foundation, 1950.
9. Fromm, Erich. *Escape from freedom*. New York: Farrar & Rinehart, 1941.
10. Getzels, J. W., and Walsh, J. J. The method of paired direct and projective questionnaires in the study of attitude structure and socialization. *Psychol. Monogr.*, 1958, 72 (1).

11. Hollingshead, A. B., and Redlich, F. C. *Social class and mental illness*. New York: Wiley, 1958.
12. Jackson, D. N., and Messick, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, 55, 243-252.
13. James, W. The true Harvard. In *Memories and studies*. New York: Longmans, Green, 1911. Pp. 348-355.
14. Kelly, G. A. *The psychology of personal constructs*. Vol. I. New York: Norton, 1955.
15. Lyle, J. G., and Gilchrist, A. A. Problems of T. A. T. interpretation and the diagnosis of delinquent trends. *Brit. J. med. Psychol.*, 1958, 31, 51-59.
16. Purcell, K. The TAT and antisocial behavior. *J. consult. Psychol.*, 1956, 20, 449-456.
17. Redl, F., and Wineman, D. *The aggressive child*. Glencoe, Illinois: Free Press, 1957.
18. Remmers, H. H., and Radler, D. H. *The American teenager*. Indianapolis, Indiana: Bobbs-Merrill, 1957.
19. Riesman, D. *The lonely crowd*. New Haven: Yale Univ. Press, 1950.
20. Rogers, C. R. A process conception of psychotherapy. *Amer. Psychologist*, 1958, 13, 142-149.
21. Sanford, N., ed. Personality development during the college years. *J. soc. Issues*, 1956, 12, 1-71.
22. Sullivan, C., Grant, Marguerite Q., and Grant, J. D. The development of interpersonal maturity: applications to delinquency. *Psychiatry*, 1957, 20, 373-385.
23. Titus, H. E., and Hollander, E. P. The California F scale in psychological research: 1950-1955. *Psychol. Bull.*, 1957, 54, 47-64.

Discussion

David V. Tiedeman, Associate Professor of Education, School of Education, Harvard University

The psychologist has had what might be termed fair success in anticipating a person's later position on some scale such as his over-all grade average in college. Even these moderate successes are often accomplished only after much investigation. Such investigations ordinarily require considerable trial and analysis of relationships existing among particular data of the past before deductions begin to square moderately well with later observations. A half century of experience of this kind has caused the modern psychologist to be highly skeptical of propositions about relationships when such propositions are not thoroughly checked out beforehand. A peculiar fascination for empiricism has been the result. This fascination sometimes spawns ludicrous claims for the value of inductive empirical study in the absence of a specific criterion as in factor analysis studies. Dr. Loevinger's awareness of such ludicrous claims probably caused her to lash out, in good humor to be sure, at the method of factor analysis as she has done today.

I could assume an air of righteousness and temporarily distract your attention from Dr. Loevinger's contribution by launching a defense of the method of factor analysis. Such action is inappropriate, though, because Dr. Loevinger gives every indication throughout her paper that she is a psychometrician *par excellence*. I prefer, therefore, to make several observations on limitations of the method of factor analysis. I shall attempt to do so by explication of the reasoning process in which Dr. Loevinger engaged.

First, let us note explicitly the several aspects of the experimental method. In essence, the experimental method consists of assembling a series of facts in which it is observed that some antecedent circumstances are associated with some consequent circumstances. A theory is then evolved which proposes that the generalized consequent circumstance has a functional dependence upon the generalized antecedent circumstance. The theory permits deductions in the form of hypotheses that some unknown, but ascertainable, consequents will be of a certain form under certain previously specified conditions of the antecedents.

These hypotheses then direct experiments in which the antecedents are created or found and the associated consequents observed. If the observed consequents agree to an important degree with the deduced consequents, we have no presumptive reason to dismiss the theory. When the observed consequents fail to agree with the deduced ones to an important degree, however, we now have a new set of antecedent and consequent observations which must be joined with our previous sets and a theory must be invented that now produces order among these enlarged data.

Now, let us turn our attention to limitations of the method of factor anaylsis. In terms of this paradigm of the experimental method it is quite apparent that the method of factor analysis aims only at the rudiments of all that is needed, namely at the introduction of some simplification of either or both the antecedent or consequent responses. Further, the method of factor analysis is usually applied with little or no intent of later investigation in mind, and hence the simplifications resulting from a factor analysis bear no necessary relationship to those that may be needed in the construction of any theory. The purpose of a theory probably offers the best guide as to the relevance of one form of simplification or another. This is not known, however, until after a factor analysis is completed unless the factor analysis is itself a purposeful step in the formulation of the theory.

We might note in this regard that the studies considered by Dr. Loevinger imply that she is more or less interested simultaneously in the processes of education, socialization and psychotherapy. By focusing her attention on these processes simultaneously, and by extracting aspects of certain studies, as well as by introducing certain observations about children, Dr. Loevinger constructs a reasonably convincing argument that we ought to consider such information in relation to an ability to conceptualize one's self. The scale she succeeds in drawing to our attention attracts because it orders, and hence gives meaning, to previously discrete information. So far no factor analysis has attempted to deal with data for such diverse circumstances and over such an extended range of age. In fact, Dr. Loevinger has amalgamated patterns of relationship *and* change that I, for one, cannot model mathematically.

There are aspects of Dr. Loevinger's approach that fascinated me, however. Remember that Dr. Loevinger is postulating an ability to conceptualize one's self, an ability generally characterized by uncontrolled impulse at its low point, separation from impulse but not from ego at the midpoint, and separation from both at the highest point. The ability is presumed to be a manifestation of ego development not related monotonically to some other aspects of ego development. It is interesting to note that Dr. Loevinger presents this scale as a Guttman-type scale, essentially postulating the absence of the type-separation from ego but not from impulse. But what if this type does appear? I have a hunch that Dr. Loevinger would deal with its appearance by designating a person of that type as "unhealthy". Thus the appearance of the type would become a cause for remedial action rather than a cause for rejection of the scale. Factor analysis, with its orientation to the patterns of responses themselves, does not allow for such judgment.

The importance of this was brought home to me several years ago when I was reflecting upon the possibility of deriving Guttman-type scales in areas of school achievement. It seemed to me that, in an area such as arithmetic, children were *expected* to master certain developmental tasks in sequence and that, as a result, a Guttman-type scale, or at least a contrived H-scale, would form around the developmental tasks. But then I began to wonder if childrens' responses in arithmetic would really scale in the Guttman sense. Although I never endeavored to answer this question, I did ask myself, "What if childrens' responses don't scale? Isn't the absence of a scaled response, *on the part of some children*, presumptive evidence that the behavior of such children is different from the intent of the teacher as scaled? Have I not gained thereby?" Actually, I believe I have, because I then possess two kinds of information: 1. what the *intended* process of differentiation and integration is, and 2. in which children the intentions are not a part of their behavior pattern and in which areas. This is diagnostic information different from an analysis of the pieces without reference to the whole, information that usually results from the direct orientation to responses characteristic of factor analysis.

In addition to the scale problem, a second fascinating aspect of Dr. Loevinger's reasoning is that, in attending to the ego, her thought necessarily turned to the *perceptions* one has of his person in relation to such notions as tasks, ideas, and people. The evaluations one holds of his person with regard to such relationships likely govern the way he orients himself to situations, and hence his behavior. I doubt that an action by itself provides sufficient information for inferring the actor's evaluation, however. The inference needs to be made from simultaneous consideration of the situation, the behavior in it, and the motivation for the behavior in it. The factor analyst has tended to limit his analyses to only one realm of activity at a time. Were motivation and behavior to be analyzed simultaneously, according to the usual model of factor analysis, I would not know if the results would uncover the effect to which I am attempting to draw your attention.

It seems to me that dynamically oriented psychologists contend that an act devoid of its motivation cannot be correctly categorized with regard to its meaning for the behavioral system of the one being studied. The factor analyst does not incorporate this judgment of meaning or function into his factor computations. Rather, the initial data for the factor analysis would include a location of subjects with respect to some categories of behavior without regard for motivation, and/or some general categorizations of degrees of motivation without much regard for situation and behavior in it. I seriously doubt that an analysis of such data would reveal factors close to the categories actually employed by an interpreter considering situation, behavior, and motivation simultaneously.

Finally, I want to note that Dr. Loevinger refers to a cognitive process, namely, ". . . a cognitive aspect of ego development." A cognitive process is understandable in terms of a succession of two subsidiary processes, differentiation and integration. Successive differentiation and integration create levels of response patterns according to the dynamically oriented psychologist. These levels are apparent only from the perspective of the growth process; I cannot see that they would be identified in a factor analysis of response patterns for a restricted range of age.

At the outset I used precious moments to sketch a paradigm of the experimental method. I then attempted to indicate: 1. that I saw no necessary reason why every categorization of information useful for psychological work need conform to responses as they exist; 2. that the character of an act may not be understood sufficiently in terms of only the act itself; and 3. that the results of differentiation and reintegration may obscure unification of relevant response because a higher level response may not appear sufficiently like its undifferentiated counterpart at a lower level. I have pursued both lines of reason because I want both to have my cake and to eat it, too.

In my judgment, Dr. Loevinger has offered no fundamental challenge of factor analysis; she has merely illustrated that it has limitations and that these limitations need to be acknowledged.

In my judgment, also, Dr. Loevinger has been scientific in her approach. She has dealt with a structure of data that factor analysis cannot analyze as her mind has done. The resulting structure she has given to an ability to conceptualize one's self is consistent with the facts she chose to consider. In addition, she has postulated relationships of this ability with other variables in ways that are subject to verification. Each of these endeavors is a part of the development of any science.

It remains to be seen whether her postulates coincide with reality to an acceptable degree. In the meantime, her present argument convinces me sufficiently so that I will be inclined to consider responses of structured personality tests in relation to her construct in the future, namely to consider the level of a response pattern in relation to a subject's age and to my judgment of his distance from impulsivity and ego.

I, for one, shall attend closely to Dr. Loevinger's investigation of this ability in the future, and would even attempt to investigate it myself if only she designates the scale more definitely than she has had time to do today.

Comment by JANE LOEVINGER

Dr. Tiedeman's use of the term "separation from" one's impulses and ego suggests pathology rather than development and evokes connotations I strove to avoid. If there is an appreciable group of people who achieve perspective with respect to their egos prior to achieving a minimal level of impulse control, the force of my argument is lost. I think of ego development as an organic growth process, a domain not preempted by Guttman. Unlike in Guttman's scale model, any specific manifestation is related to ego development only probabilistically; the more specific, the lower the probability. Hence a crucial test of my hypothesis is not altogether simple.

Measurement and Prediction of Teacher Effectiveness

DAVID G. RYANS, Department of Educational Psychology, The University of Texas

Proper consideration of even some of the more obvious problems associated with the definition, measurement, and prediction of teacher effectiveness would be, as you all well know, an exhausting undertaking. What I shall attempt is to sketchily review with you a few of the complexities facing researchers who try to study teaching competency; then go on and briefly summarize some high spots in research conducted by the Teacher Characteristics Study; and finally, suggest some tentative conclusions about identifiable conditions and characteristics which may be associated with teacher effectiveness.

Some Basic Issues

The basic concern of research on teacher effectiveness is, of course, prediction. We seek to determine how and to what extent various data descriptive of teachers (e.g., verbal responses, overt acts, biographical information, etc., all of which may be subsumed under *teacher characteristics*) are either 1. antecedents or 2. concomitants of some behavior agreed to be a component of some criterion of teaching competence.

The extent to which such relationships can be uncovered depends, of course, not only on the real, or latent, relationships which may obtain, but also on 1. how unambiguously and operationally the *agreed upon* criterion can be defined, and how validly and reliably estimates of the criterion can be obtained, 2. how unambiguously the teacher characteristic under study can be identified and how validly and reliably it can be measured, and 3. what the purposes and hypotheses of the research are and how adequately it has been designed, taking into account sampling, control, and replication. I should like to deal briefly with these three areas of problems.

Criterion Measurement. Recently I attempted to outline different methods of obtaining criterion data relative to teacher effectiveness.

The major categories included: a) direct measurement of *on-going* teacher behavior (e.g., time sampling involving replicated systematic observation); b) indirect measurement based on preserved records of on-going teacher behavior (e.g., tape recordings); c) indirect measurement by non-trained observers, based on recall of teacher behavior and assessment thereof (e.g., ratings by students, administrators, peers, etc.); d) measurement of a product (student behavior) of teacher behavior; and e) measurement of concomitants (secondary criterion data) of the criterion of teacher effectiveness.

These different approaches to measurement vary in nature of rationale employed to support them, in reliability of the criterion data produced, and in the order of obtained relationships between criterion estimates, thus differently derived, and specified predictors—this last observation, of course, merely bearing testimony to the fact that most criteria are very complex and any one set of estimates is likely to be very incomplete with respect to the overall criterion.

Approaches to the measurement of a criterion of teacher effectiveness thus involve the evaluation of either 1. teacher behavior *in process*, 2. a *product* of teacher behavior, or 3. *concomitants* of teacher behavior. Measurement of on-going behavior of the teacher is the most direct approach; measurement of products and of concomitants are less direct and more subject to the effects of confounding conditions.

Concomitants (which, in a sense, may be thought of as secondary criterion data) usually are not acceptable for criterion measurement when direct measurement of behavior in process or the measurement of isolable products of teacher behavior can conveniently be used. However, in investigations involving extensive sampling and where other measurement approaches are impractical, the use of known correlates as substitutes for process or product data frequently is defensible.

Of the measurement approaches employing observation and assessment only *time sampling involving replicated systematic observation by trained observers* produces sufficiently reliable data to recommend its use in fundamental research, although less well-controlled variations (e.g., ratings by students) may be employed when only coarse discrimination (e.g., "best" and "poorest" teachers with respect to some criterion component) is required, and when the larger expected error is recognized and accepted. Various assessment techniques have been developed, among which the more reliable and promising appear to be 1. graphic scales with operationally, or behaviorally, defined poles and/or units, 2. observation check lists, and 3. forced-choice scales. The chief shortcoming of observation and assessment techniques has been lack of reliability, a shortcoming which research has indicated can fairly readily be overcome with care to definition and to scale development, and with adequate training of the observers or judges.

Product measurements (estimates of the behavior or achievement of the pupils of teachers) have been widely acclaimed as desirable criterion data, but have been infrequently used in the study of teacher effectiveness. Actually, the seeming relevance and appropriateness of the measurement of pupil behaviors and their products as indicators of teacher performance may be more apparent than real, for the producers of (or contributers to) pupil behavior, or pupil achievement, are numerous, and it is most difficult to designate and parcel out the contribution to a particular "product" made by a specified aspect of the producing situation, such as the teacher. We also must note that the facets of the product criterion (various understandings, skills, and attitudes, etc. in various content fields and areas of personal behavior) are similarly numerous, and each must be capable of valid measurement and of at least partial isolation for study. The comparability of estimates of various components or aspects of a product (pupil achievement, for example) also becomes a special problem when student behavior or achievement is employed as a criterion of teacher effectiveness. And when measurement of the product is accomplished by obtaining estimates of student change (i.e., pretest-posttest data) the problem of variable potential gain (students who score high on the initial measurement being closer to their "ceilings" than students who originally score low are to theirs) is particularly plaguing to the researcher. However, if the rationale of the product (student performance) criterion is accepted, and if the complex control problem presented by a multiplicity of producers and the multidimensionality of the criterion can be satisfactorily coped with, student change becomes an intriguing approach to the measurement of teacher effectiveness.

In dealing with any of the several approaches to measuring the criterion, the researcher must be thoroughly familiar with, and guard against, the various sources of criterion measurement bias, particularly those which have to do with a) incompleteness and b) contamination of the obtained data.

Predictor Measurement. I shall pass very quickly over the problem of obtaining estimates of the predictors. The chief technical problems faced here are those familiar to educational research and measurement workers, namely validity and reliability. The prediction of a criterion may be very seriously limited by the reliability of estimates of the predictor employed in a study. And unless the researcher has a pretty clear idea of the meaning of his predictor estimates and the conditions or traits they actually represent, interpretation of predictor-criterion relationships may be pretty risky.

It is important to note that similarly *named* predictor measures (e.g., estimates of teacher empathy, or leadership, or understanding of children) used in different investigations do not necessarily refer to the same underlying characteristic of the teacher which is measured. Quite apart from sampling errors, they do not necessarily yield similar relationships with estimates of a specified criterion dimension. Discrepancies in findings reported in the literature sometimes may be traced to this lack of agreement in operational definition of the predictor, in addition to criterion inadequacies and lack of control of relevant variables.

Research Objectives and Design. Still another set of conditions which contribute to variability in the nature and degree of association which may be obtained between hypothesized predictor measures and measures of a criterion of teacher effectiveness has to do with research objectives and the approach to the predictor-criterion relationship incorporated in the research design. Such questions as the following should be (but frequently are not) considered by the researcher.

1. Does the investigation purpose to determine a) concomitant or b) antecedent-consequent relationships?
2. Is prediction of the criterion of teacher effectiveness attempted from single bits of information (e.g., answers to a single questionnaire, test, or inventory item) or from scores based on combinations of such bits of information forming sets of homogeneous items, or scales? (And, if the latter, does the combination of bits involve equal or differential weighting?) An extension of this question involves whether prediction of the criterion is determined from a single predictor alone or from a combination of predictor scores, weighted perhaps in light of multiple regression weights.
3. Is the derivation of predictors (original selection of items, or combinations of items, as predictors of the criterion) based upon experience with a single sample, or has replication been employed involving multiple samples of teachers?
4. Is prediction directed at a) additional random samples of the same population as the samples employed in deriving the predictors (e.g., cross validation) or b) samples of populations other than that from which the predictors were derived either 1. employing the same criterion measure (validity generalization) or 2. a different criterion measure (validity extension)?
5. Is prediction attempted for predictor data and criterion data which have been collected at approximately the same time, or when the

obtaining of criterion data has been delayed and carried out with a considerable time interval separating the collection of the two sets of estimates?

6. Is prediction attempted when the predictor data are obtained under "incentive" conditions (e.g., in connection with selection for employment) or under "non-incentive" conditions (e.g., as in basic research)?
7. Is prediction attempted for selected criterion dimensions singly (e.g., effective classroom discipline) or for a composite criterion made up of a number of heterogeneous components or dimensions (e.g., overall teaching effectiveness)?
8. Is prediction of teacher effectiveness attempted on an actuarial, or group, basis, or is the concern prediction for particular (individual) teachers?

Still other aspects of the prediction problem might be noted, but these are representative of some of the major considerations involved in the overall design of studies of the predictor-criterion relationship.

Teacher Effectiveness and the Teacher Characteristics Study

When, near the beginning of this discussion, I referred to methods of obtaining criterion data relative to teacher effectiveness, I avoided definition of the term "teacher effectiveness." If I were pressed I might say that I believe teaching is effective to the extent that the teacher acts in ways that are favorable to the development of basic skills, understandings, work habits, desirable attitudes, value judgments, and adequate personal adjustment of the pupil. But even such an operational-appearing definition really is very general and abstract and is not easily translatable into terms relating to specific teacher behaviors. Embarrassing as it may be for professional educators to recognize, relatively little progress has been made in rounding out this definition with the details which are necessary for describing competent teaching or the characteristics of effective teachers for a specific educational situation or cultural setting. Granted, most educators and most parents do have some idea of what constitutes effective teaching. These conceptualizations, however, usually are very vague and far removed from specific observable behaviors of teachers. Frequently even such hazy ideas are highly individualized with very little agreement existing among different persons. One is reminded of the old, familiar fable of the blind men who perceived an elephant in widely varying manners depending on the part of the elephant's body that each one touched.

Relativity of Teacher Effectiveness. Disagreement and ambiguity with respect to the description of teacher effectiveness are to be expected and cannot be entirely avoided because competent teaching undoubtedly is a relative matter. A person's concept of a "good" teacher depends, *first*, on his acculturation, his past experience, and the value attitudes he has come to accept, and, *second*, on the aspects of teaching which may be foremost in his consideration at a given time. Pupil F, therefore, may differ widely from pupil G in his concept of the essential attributes of an effective teacher. If pupil F is bright, academically minded, well adjusted and independent, he may value most the teacher who is serious, rigorously academic, and perhaps relatively impersonal. If pupil G, on the other hand, is more sensitive and requires considerable succorance, he may find the teacher just described not at all to his liking and indeed literally "impossible." In the mind of pupil G, the better teacher may very well be one who is somewhat less exacting from an academic standpoint, but who is characteristically sympathetic, understanding, and the like.

Answers to the question, "What is an effective teacher like?" also may vary to a degree with the particular kind of a teacher one chooses to consider. It does not seem unreasonable to hypothesize that, even if it were possible to agree upon a generalized definition of effective teaching which would be acceptable to a number of different cultures, and if our thinking might be objectified to the point where effective teaching could be described on a factual basis, "good" teachers of different grades or different subject matters still might vary considerably in personal and social characteristics and in various domains of classroom behavior.

The concept of competent teaching must therefore be considered to be relative to at least two major sets of conditions: 1. the social or cultural group in which the teacher operates, involving social values which frequently differ from person to person, community to community, culture to culture, and time to time, and 2. the grade level and subject matter taught. It is not suprising, then, to note the difficulties that have confronted those seeking to establish criteria of teacher effectiveness, the dearth of testable hypotheses produced in such research as has been undertaken, and a general lack of understanding of the problem of the characteristics of effective teachers. One very important reason why effective or ineffective teachers cannot be described with any assurance is the wide variation in tasks performed by the teachers and in value concepts of what constitutes desirable teaching objectives.

But in addition to these considerations, and important in its own right as a deterrent to the study of teacher effectiveness, is the fact that there is a lack of any clear knowledge of the *patterns of behaviors* that typify individuals who are employed as teachers. It seems probable that, without losing sight of the importance of developing means for recognizing "good" teachers, attention of the researcher might first more properly and profitably be directed at the identification and estimation of some of the major patterns of personal and social characteristics of teachers. This represents the point of departure for research conducted by the Teacher Characteristics Study.

In the Teacher Characteristics Study, considerations of the effectiveness, or value, of particular teacher behaviors were to a large extent disregarded. Instead, attention was focused on the study of possible teacher behavior dimensions, such dimensions being hypothesized to represent generalized trait continua. From this point of view teacher behavior variables are assumed to consist of clusters of relatively homogeneous (positively intercorrelated) behaviors, such component behaviors being of the nature of *simple predicates*, capable of operational definition.

Implied in this approach is the assumption that a teacher may be described in terms of a position on a particular behavior dimension, such description being essentially factual and relating to observable manifestations of overt behavior or else to responses known to be correlated with some behavior pattern to a degree that may permit indirect estimation of the behavior.

The Teacher Characteristics Study. The Teacher Characteristics Study was sponsored by the American Council on Education and generously supported by The Grant Foundation. During the six years of the Study approximately 100 separate research projects were carried out and over 6000 teachers in 1700 schools and about 450 school systems participated in various phases of the research. Some of the basic studies involved extensive classroom observation (by trained observers) of teachers, with the purpose of discovering significant patterns of teacher behavior. Other activities of the project had to do with the development of instruments (paper and pencil tests and inventories) for the identification of individuals characterized by different levels of specified patterns of a) classroom behavior, b) attitudes and educational viewpoints, c) verbal intelligence, and d) emotional stability. Still other investigations were concerned with the comparison of defined groups of teachers (e.g., elementary teachers and secondary teachers, married and unmarried teachers, etc.), from the standpoint of their observable characteristics.

Basically, the Teacher Characteristics Study had three major purposes: 1. to analyze and describe patterns of teacher classroom behavior and the manifestations of certain value systems and cognitive and emotional traits of teachers; 2. to isolate and combine into scales significant correlates (provided by responses to self-report inventories concerned with the teacher's preferences, experiences, self appraisals, judgment, and the like) of some major dimensions of teacher behavior; and 3. to compare American teachers (in terms of the teacher characteristics described by the Study) when they had been classified according to a number of conditions.

Pursuance of these objectives involved development of techniques for the reliable assessment of classroom behavior, determination (largely through factor analysis) of some major patterns of teacher behavior, development of instruments made up of materials hypothetically related to teacher classroom behavior dimensions and other personal and social characteristics of teachers, the empirical derivation of scoring keys for such instruments in light of response-criterion correlations, and finally comparison of defined groups of teachers.

Patterns of Classroom Behavior. As a result of the direct observation and assessment of teacher classroom behavior and subsequent statistical analyses of the measurement data, several interdependent patterns of teacher behavior were suggested. Three in particular appeared to stand out in separate factor analyses of elementary and secondary teachers:

T.C.S. Pattern X_o—understanding, friendly *vs.* aloof, egocentric restricted teacher behavior

T.C.S. Pattern Y_o—responsible, businesslike, systematic *vs.* evading, unplanned, slipshod teacher behavior

T.C.S. Pattern Z_o—stimulating, imaginative, surgent *vs.* dull, routine teacher behavior

Pattern scores X_o, Y_o, and Z_o, derived from observers' estimates of teacher behaviors in the classroom, appeared to possess sufficient reliability to permit comparisons of teacher groups with respect to these patterns and, also, to justify their use for criterion purposes in attempting to identify inventory responses which might be used to predict teacher classroom behavior.

Among elementary school teachers, patterns X_o, Y_o, and Z_o were highly intercorrelated and each also seemed to be highly correlated with pupil behavior in the teachers' classes. Among secondary school teachers the intercorrelations of the patterns were less high, than between patterns X_o (friendly) and Y_o (organized) being of a very low order. The teacher classroom behavior patterns and pupil behavior were much less highly correlated among secondary teachers as compared with elementary teachers.

Elementary and secondary teachers, as major groups, differed hardly at all with respect to mean assessments on patterns X_o, Y_o, and Z_o. However, grade 5-6 women teachers, represented by a relatively small sample, were assessed somewhat higher on the several classroom behavior patterns (particularly on Y_o) than teachers of other elementary grades. Among secondary school groups, social studies teachers received the highest mean assessment on pattern X_o (friendly behavior) and women mathematics teachers (with women social studies teachers not far behind) on pattern Y_o (businesslike behavior). Teachers over 55 years of age received distinctly less high mean assessments on pattern X_o (friendly), and also slightly lower with regard to pattern Z_o (stimulating), than younger teacher groups. Among elementary teachers the mean assessments on the classroom behavior patterns X_o, Y_o, and Z_o were slightly but insignificantly higher for married as compared with single teachers. Among secondary mathematics-science teachers, single teachers received higher mean assessments than did those who were married. With respect to English-social studies teachers, single teachers were assessed higher than married teachers on pattern Y_o, but slightly lower on patterns X_o and Z_o. In general, differences between teacher groups compared on the observed classroom behavior patterns X_o, Y_o, and Z_o were not pronounced. However, it is of interest to note that scores on the Teacher Characteristics Schedule (to be described shortly), based on keys (X_{co}, Y_{co}, and Z_{co}) derived to predict these classroom behavior patterns, frequently distinguished different teacher groups more sharply and with greater assurance than did the X_o, Y_o, and Z_o observation data.

Patterns of Values, Verbal Ability, and Emotional Stability. Inevitably the Teacher Characteristics Study sought other evidences of teacher behavior in addition to those provided by assessments of overt classroom behavior. To extend the understanding of conative and cognitive aspects of teacher behavior, and to permit the more complete investigation of relationships between teacher characteristics and specified conditions of teaching, the study undertook a number of researches directed at analyses of teacher's attitudes, their educational viewpoints, their verbal intelligence, and their emotional adjustment, and attempted to develop direct inquiry type instruments for estimating from a teacher's responses his status relative to such behavior domains.

In one set of studies a number of opinionnaires relating to teachers' attitudes toward groups of persons contacted in the school were developed and the organization of teacher attitudes was studied through factor analysis. In keeping with the results of the factor analyses the study centered its attention chiefly on the attitudes of teachers toward pupils, their attitudes toward administrators, and their attitudes toward fellow teachers and non-administrative personnel.

The educational viewpoints of teachers with respect to curricular organization and scope, pupil participation and class planning, academic achievement standards, etc. also were investigated (separately for elementary and secondary teachers) through the employment of direct inquiry type items and factor analysis of the intercorrelations among responses. The patterns of viewpoints which emerged were not clear-cut and there seemed to be some justification for considering teachers' educational beliefs from the standpoint of a single continuum, over simplified perhaps by its designation as a "traditional-permissive" dimension.

To obtain estimates of the verbal understanding of teachers, vocabulary and verbal analogy items were constructed, experimentally administered, and the responses analyzed, the procedure culminating in the selection of a small number of highly discriminating items comprising a "verbal ability" scale. In a similar way materials were prepared and analyzed to obtain items for providing estimates of the emotional stability of teachers. And to aid in the detection of "tendency to make a good impression" when dealing with responses to direct question type materials, a set of items intended to measure probable validity-of-response of teachers also was assembled.

Various studies and comparisons of the attitudes, educational viewpoints, verbal understanding, and emotional adjustment of teachers were undertaken in the course of the development of such measuring devices as those noted above. Some of these results were extremely interesting, but I shall not attempt to go into them here. I shall move on to a description of our efforts to obtain indirect estimates of teacher classroom behaviors and other characteristics from correlated inventory responses.

An Inventory for Indirect Estimation. In the interest of providing more readily obtainable estimates of teacher classroom behaviors, and also estimates of teacher attitudes, viewpoints, verbal ability, and emotional stability which might be less susceptible to the response set of giving socially acceptable responses, efforts of the Teacher Characteristics Study were directed at the derivation of correlates scoring keys applicable to the items of the Teacher Characteristics Schedule. The Teacher Characteristics Schedule was an omnibus self-report type inventory based upon some 25 originally separate instruments. In its final form, it consisted of 300 multiple choice and check list type items relating to personal preferences, self-judgments, frequently engaged in activities, biographical data, and the like.

Employing as criteria a) observers' assessments of teacher classroom behaviors X_o, Y_o, and Z_o and b) scores on the direct response scales relative to teacher attitudes, viewpoints, verbal intelligence, and emotional stability, hundreds of response analyses were carried out (thanks to SWAC, our first high speed computer at UCLA). Response-criterion correlations were obtained for each response to each item of the Teacher Characteristics Schedule under a variety of conditions. Correlates scoring keys, employing responses associated with the criterion behaviors as signs or symptoms of behavior, thus were derived for a large number of teacher groups. The most generally applicable sets of scoring keys (and those most frequently used in other phases of the study's research) were the all-elementary teacher keys, the all-secondary teacher keys, and the combined elementary-secondary teacher keys.

Reliability data for the correlates scoring keys and various kinds of validity data, relating particularly to the friendly (X), business-like (Y), and stimulating (Z) keys were obtained. Generally speaking, the reliability coefficients fell between .7 and .8 and the validity coefficients were of varying magnitude depending upon the kind of validity investigated, the particular behavior estimated, and the teacher group from which the key was derived and to which it might reasonably be applied. Concurrent validity coefficients for correlates scores on classroom behavior patterns X, Y, and Z typically were between .2 and .4; predictive validity coefficients were positive, but generally low, seldom exceeding .2. Intercorrelations among scores resulting from application of the several correlates scoring keys estimating classroom behaviors, attitudes, educational viewpoints, verbal intelligence and emotional stability, and correlations between Schedule scores and observers' assessments, indicated 1. substantial relationships among the correlates data and 2. prediction of observed classroom behaviors principally by the scales specifically developed for that purpose (X_{co}, Y_{co}, and Z_{co}).

"High" and "Low" Teachers Compared. I shall not deal here with the numerous comparisons of teachers which were made in light of the Teacher Characteristics Schedule data collected. But I do want to mention a study we conducted which was concerned with identifying teachers who fell into one of three groups: one group comprised of teachers each of whom had received observer assessments one standard deviation or more above the mean on each of the three classroom behavior patterns X_o,

Y_o, and Z_o; another made up of teachers who were all within two-tenths of a standard deviation on either side of the mean on the three different classroom behavior patterns; and a third group made up of teachers all of whom received observers' assessments one standard deviation or more below the mean on each of the three classroom behavior patterns. After having identified these teachers we attempted to determine some of the distinguishing characteristics, in terms of Teacher Characteristics Schedule responses, of the different groups. Here, I suppose, we were approaching the problem of over-all teacher effectiveness. We were attempting to discover responses of generally highly assessed teachers which distinguished them from generally lowly assessed teachers. I shall summarize some of the more notable characteristics, for combined elementary and secondary teachers, which distinguished the high group from the low and the low group from the high. There was a general tendency for "high" teachers to: be extremely generous in appraisals of the behavior and motives of other persons; possess strong interest in reading and literary affairs; be interested in music, painting, and the arts in general; participate in social groups; enjoy pupil relationships; prefer non-directive classroom procedures; manifest superior verbal intelligence; and be above average in emotional adjustment. Turning to the other side of the coin, "low" teachers tended generally to: be restrictive and critical in their appraisals of other persons; prefer activities which did not involve close personal contacts; express less favorable opinions of pupils; manifest less high verbal intelligence; show less satisfactory emotional adjustment; and represent older age groups.

Obviously, the description I have been able to give of the Teacher Characteristics Study is very sketchy. I have not been able to get down to some of the really very interesting findings such as those related to comparisons of teacher groups and interrelationships among teacher behaviors. I will, however, be able to incorporate some of our findings in the concluding section which follows.

Some Probable Correlates of Teacher Effectiveness

It is indeed presumptuous and dangerous to speak out boldly about conditions and teacher characteristics associated with teacher effectiveness. However, based upon the findings of various researches conducted by the Teacher Characteristics Study and an accumulation of investigations which have appeared in the literature over a period of years, certain threads of fact do seem discernible. But the conclusions and inferences

65

are still, at best, tentative—they are more in the nature of hypotheses for which some support has been found in our American midtwentieth century culture. And we also must recognize that changing educational values in the future reasonably may lead to changes in the patterning of teacher behaviors and to further revision of our understanding of predictors of teacher effectiveness.

The following generalizations regarding the relationship between teacher characteristics, as predictors, and teacher effectiveness, as a criterion abstracted from various criterion measures reported in the literature, appear to be in order.

Characteristics and conditions of the teacher which are likely to be *positively correlated* or associated with teacher effectiveness in the abstract include: 1. measured intellectual abilities, particularly verbal intelligence; achievement in college courses; general cultural and specific subject matter knowledge; professional information (knowledge of education and teaching); practice teaching marks; emotional adjustment; attitudes favorable to students or pupils; generosity in appraisals of the behaviors and motives of other persons; interest in reading and literary matters; interest in music and painting; participation in social and community affairs, early experiences in caring for children and in teaching (e.g., reading to children, taking class for teacher), history of teaching in family, size of school and size of community in which presently teaching, and cultural level of community in which teaching. 2. Extensiveness of general and/or professional education, enrollment in particular professional courses, and personal appearance appear to bear very little relation to the abstracted criterion of general teacher effectiveness. 3. Elementary teachers and secondary teachers, as groups, do not seem to differ greatly when an over-all view of teacher effectiveness is taken. However, elementary teachers do seem to show superiority when selected aspects of criterion behavior having to do with warmth, permissiveness, and favorable attitudes toward children are considered. Secondary teachers are superior from the standpoint of verbal understanding. Within the elementary school, Grade 5-6 teachers tend toward superiority on several criterion dimensions; within the secondary school, English and social studies teachers show a similar tendency. 4. Age of the teacher and amount of teaching experience seem to manifest an over-all negative relationship with teaching effectiveness, although there is evidence of curvilinearity, increase in effectiveness appearing to be positively correlated with age and experience during early years of teaching careers. 5. Sex differences in over-all teacher effectiveness do not appear to be pronounced, but the classroom performance of women teachers seems to be more organized and businesslike than that of men, and men teachers seem to be very distinctly more emotionally stable. 6. For teachers of all grades and subjects considered together differences in effectiveness between single and married teachers are small. However, within the elementary school the evidence appears to favor married teachers. At the secondary level results are somewhat mixed, with unmarried teachers as a group appearing to be superior with respect to such criteria as business-like classroom behavior, permissive viewpoints, and verbal understanding, but with married teachers showing superior emotional adjustment.

Certain characteristics, then, do seem to be associated with certain dimensions of teacher behavior and teacher effectiveness, although the extent of obtained relationships frequently has not been high. It is important here to recall that 1. relationships and differences which have been noted are in terms of averages for groups of teachers, and 2. any obtained relationship is limited by, and may be expected to vary with, conditions such as those noted in an early part of the paper. The usefulness of research findings pertaining to the prediction of teacher effectiveness will be greatest when the results are considered in an actuarial context, rather than in attempting highly accurate prediction for given individuals, and when variations in relationship found among different classifications of teachers, and with use of different approaches to the predictor-criterion relationship, are taken into account.

5

Appendix: Predictability of Teacher Effectiveness

The notes which follow have to do with general considerations relating to conditions which probably should be taken into account both in the design and the interpretation of research on teacher effectiveness. Some of these are derived from rational analysis of the problems involved, but many also have substantial support from empirical data.

1. The predictability of teacher effectiveness undoubtedly is affected by the *multi-dimensionality of the criterion*. There is accumulating evidence that prediction can be accomplished with better than chance results for specified dimensions or components of the criterion. On the other hand, the prediction of over-all teacher effectiveness is possible only to the extent that some general agreement can be reached regarding the dimensions comprising "over-all effectiveness" (involving, of course, acceptance of a common set of educational values) and how they should be combined to form a composite. Teachers effective with regard to one

aspect of the criterion may not be effective when judged by other criterion dimensions.

2. The predictability of teacher effectiveness varies depending on the *degree of control* it is possible to exert in dealing with the multiplicity of predictors and the multidimensionality of the criterion.

3. The predictability of the criterion varies with the *kind of measure* employed in obtaining the criterion data.

4. The predictability of the criterion varies with the *adequacy* (reliability and validity) *of measures* of a) the criterion and b) the predictor variables.

5. The predictability of the criterion is so limited by conditions associated with measurement of the criterion, measurement of predictors, and practical conditions, that relationships representing common variance of perhaps one-fifth or one-fourth of the total variance probably approach the maximum to be expected except in chance instances.

6. The predictability of a dimension of the criterion of teacher effectiveness from a specified predictor probably varies depending upon the cultural milieu which provides the setting for an investigation, particularly the values and objectives prominent in the teacher training curriculum at the time the teachers studied were in college.

7. Predictability of the criterion varies directly with the degree of similarity between the sample with respect to which predictors are derived, and the sample to which the predictors are applied in attempting to determine predictor-criterion relationships.

8. Predictability of a criterion dimension varies with the *particular teacher population* (e.g., Grade 1-2 women teachers, men science teachers, etc.), or student population, studied. Effective teaching methods may differ from one grade level to another and from course to course.

9. Predictability of the criterion varies inversely with the *time interval* separating the obtaining of predictor measurements and criterion measurements.

10. Predictability of the criterion probably varies depending upon the association of *incentive or non-incentive conditions* with the obtaining of predictor data.

11. The regression of predictor measurements on criterion measurements frequently is *curvilinear* (e.g., positive correlation between amount of teaching experience and certain criterion measures of effectiveness of secondary school teachers during first five years or so, followed by leveling off and decline in criterion estimates with extensive experience).

12. Prediction of teacher effectiveness must be considered largely in the *actuarial* sense; individual prediction, as generally is the case in

attempting to predict human behavior, is much more limited and is accomplished with a lesser degree of confidence.

Discussion

HARRY B. GILBERT, Board of Examiners, New York City Board of Education

The role of a discussant in following Dr. Ryans' paper is unenviable, in many respects, but a distinct privilege in others. First let us examine the unenviable aspects. Dr. Ryans has given us a splendidly condensed version of the compressed results of a series of 100 separate research studies ranging from elaborate to complex—studies extending over a period of years, involving thousands of teachers and at least 1700 schools in 450 school systems. Does the discussant proceed to address himself to problems of methodology, techniques, inferences from the data and the like? I think you must share my feeling that I should not, even if I were able to. The scope is too vast to be treated thus.

I am reminded of a sign that greets me regularly in my favorite coffee shop. It reads "Use your head—it's the little things that count" and that is precisely my estimate of self in relation to the magnitude of the studies under consideration. I have a deep respect for Ryans and his colleagues, and the subject they have been studying, and, I hope, a proper appreciation of little old me in that big old context.

I have therefore chosen to discuss some general implications of Ryans' studies, and here I feel deeply privileged for the opportunity.

Surely there is no need to dwell overlong on the social significance of the studies. All of us are aware of the shortage of teachers. We are aware of the shrinking supply of future teachers in our colleges, and of our increasing pupil population. It spells trouble now and deeper trouble ahead for all of us who see education as the country's pressing concern and basic approach to the development of a sound nation and world.

Therefore, for those of us who are directly involved in the selection of teachers we look with great interest to all research that can be of help to us. We in the New York City school system are in big business in teacher selection. We examine about 30,000 applications annually and need replacements at all levels numbering in the thousands each year. We have addressed ourselves to the problem of determining scopes of examinations, constructing tests, setting pass marks in advance, somewhat uneasily aware of the tremendous complexity of devising prediction instruments for unestablished and undefined criteria.

Ryans' results give us much cause for concern. He points up the difficulties in determining teacher effectiveness, particularly in attempting to compress the many dimensions into a single over-all categorization. Yet we go right on making our tests designed to predict overall teaching success.

Another point. Only recently, the Superintendent of Schools in this city has gone on record favoring an advance in salary for teachers of "superior merit." There are many school systems throughout the country that have such a policy in effect. However, if there is one conclusion we can draw from the studies under discussion it is that the plain facts indicate we cannot select such "superior merit teachers" with any degree of confidence. The idea of superior merit is intriguing. However, the results of objective investigation should give pause to school administrators, entirely apart from the host of other arguments that can be brought to bear against the adoption of such a proposal. I refer, of course, to the negative morale factors, the inevitable rivalries among teachers, the invitation to currying favor and the like.

But all is not black. One notes with a sigh of relief some of the comparisons of "high" and "low" teachers that are reported. I call to your attention that "high" teachers have been found to be superior in verbal intelligence, interest in reading and above average in emotional adjustment in contrast with "low" teachers who are inferior in these dimensions. I have been deliberately selective in citing these characteristics. Our selection procedures are loaded with these variables. This is not the time to discuss the validity or reliability of the instruments we use. It is of importance to be reassured, for whatever it is worth, that our selection procedures are designed in what appears to be some desirable dimensions.

Certainly the conclusion—the very obvious conclusion—we must make is that the studies should invigorate us with the determination to continue the research that has been started. It should make us feel very humble about our too-entrenched, too-established procedures and fill us with the conviction that we must not be floored by the complexity of the problem. We have to learn the dimensions of teacher effectiveness and how to predict them. This may mean an entirely new approach to selection, but at least we must proceed in the directions that the research leads, and not, as we have been doing too frequently, proceed in the directions of test construction in which we are most competent or in just plain interviewing, which our egos will not permit us to question.

The current stage is most opportune for extension of this research. It is probably true that most people who complete teacher training can find jobs as teachers. They may not all be placed in the precise locale they desire. But the shortage is such that jobs can be found. Note the difference from the situation in the 1930's, for example. In that decade, the job of the teacher was highly sought. There was wholesale attrition among applicants. Any research then on prediction of teacher effectiveness was severely limited because of the large turn-away.

The situation, I repeat, is radically different today and in the near foreseeable future. Hence it would be most urgent to institute a crash program of significant dimensions to extend and refine Ryans' work. Witness the current struggles between the liberal arts and teachers' college advocates. Note the presumptuous assertions by laymen regarding education and the preparation of teachers.

The trouble, as ever, is that those who proceed cautiously and are aware of the complexity of the problem are least likely to make their voices heard.

I end with the note that we have a public responsibility to make known our need to undertake a vast program of research, designed to provide fundamental guides to teacher training, teacher selection and inevitably to in-service training and supervision of teachers. We are indebted to Ryans and his colleagues, but we shall let them down, and let ourselves down, if we do not insist on the logical continuation of the work.

Luncheon Address

Some Observations on Soviet Education

by **Henry Chauncey**, President, Educational Testing Service

A number of years ago, when I was talking with General Hershey in his office, he suddenly stopped, turned to me and said, "You are probably like me—just have one speech. Sometimes I begin at the beginning and go to the end; sometimes I begin at the end and go to the beginning; sometimes I begin in the middle and go both ways. That confuses them, but they think it's profound."

I used that story to introduce General Hershey at an Invitational Conference that some of you may remember, and I use it again today. For certainly, with regard to Russia, I have but one speech, and I am afraid that I may seem to enter it in the middle and go in several directions. I hope, however, that it will not be confusing and I hope further that you will not think it profound, since it represents merely the observations of one visitor to the Soviet Union. I happened to have the good fortune to be a member of the first American educational team to visit Russia under the Cultural Exchange Agreement. The team went over under the auspices of the Office of Education and was led by Lawrence Derthick, the United States Commissioner of Education. It was our assignment to look into elementary and secondary education and also teacher training.

We spent a month in the Soviet Union and traveled fairly widely. After a week in Moscow, we went to Kazan, the capital of the Tartar Republic; Sverdlovsk, in Siberia; Alma Ata in Khazakstan, near the Mongolian border; Tashkent, in Uzbekistan, across from Afghanistan; to Sochi on the Black Sea; then to Minsk in Belorussia, and Leningrad, then back to Moscow for final visits with various people in the Ministry of Education.

Our hosts modified the itinerary and plans they had made for us to suit our wishes and gave us opportunity, within practical limits, to see whatever we wanted to see. While I feel sure that we did not visit some of the worst schools in the Soviet Union, we saw a broad range of institutions. Since in each city the team split up into smaller groups, we were able to visit a large number of institutions—well over a hundred.

I think I might begin by indicating the three things that struck me most about education in the Soviet Union. First, and most important of all, is the tremendous commitment of the people of the country to education. Secondly, the flexibility of the country in educational matters, as in other ways. And, thirdly, the great progress that has been made in education during the last forty years.

The foremost impression that our whole team had was the strong commitment of the Soviet people to education, their conviction that this is tremendously important, and their desire to do everything to improve education and strengthen it.

Now, the reason for this is that they believe education is the foundation of national power. They believe that if they are going to be strong, scientifically, militarily, economically, they must be an educated country, educated in many different ways, and that this is the basis on which they will grow from power to power.

They are a country with a tremendous power drive, and their aim, as they express it on bulletin boards all over the Soviet Union, is to "reach and surpass America." This is their goal, and education is the foundation stone.

A second quality, flexibility, that struck me forcibly also came as a surprise. One thinks of Russia as a dictatorship, a monolithic enterprise that is moving ahead relentlessly in one direction, a direction that will later prove to be wrong. But the fact of the matter is that the Russians are tremendously flexible and adaptable. They move ahead, but if they find conditions require a different tack, they make it.

When I talked to Henry Shapiro, the UP correspondent who has been there 25 years, his comment was that Russia is the most flexible country in the world. People just don't understand or recognize this, but the Soviet Union is continually changing and adapting, moving ahead, moving aside, as conditions make necessary. This is true in education as well as in many other ways.

The third observation that I want to mention particularly is the great progress the Russians have made in their education over the last 40 years. One could not but be impressed by this, because they talk about it all the time; nevertheless, they have facts and figures to back it up.

They started with a large number of obstacles to overcome: a tremendous country, sparsely populated, with a very small educational system that was intended particularly for the nobility, for the elite. They were 70 percent illiterate in 1917. They didn't have an industrial economy to back up any progress that they might plan.

And then in the midst of their development during the last 40 years, they had what they call "the Great Patriotic War," and the whole western half of their country was overrun and devastated. Despite this, they have continued steadily to make progress in the field of education.

Let me give you an example from one Republic, because I think specifics usually make the situation a little clearer. In the Republic of Uzbekistan, down near Afghanistan, the population was 98 percent illiterate in 1917. There were 160 schools with 17,300 pupils. Today illiteracy has been virtually wiped out in Uzbekistan. There are 5,800 schools with 1,300,000 students. In 1917 there were no institutions of higher education; today there are 34.

Education in the Soviet Union has gone through a number of phases since the 1917 Revolution. It is useful in understanding Russian education today to recapitulate briefly the steps that have been taken since the Communists came into power. Before the Revolution, the Russians had a European type of educational program. It was strictly academic, very rigorous, and intended for the intellectual elite, not for people of the country as a whole. After the Revolution, the Communists threw the whole system out and decided to have education for everybody along "progressive" lines, then in vogue in some countries.

But somehow they carried it a little too far, and in the early 30's they began to be disillusioned by progressive education. They found that it was not producing the kind of trained individuals for scientific work or for other kinds of leadership that they needed. So, with their customary facility for making an about-face, they discarded progressive education completely—and with it, incidentally, all use of objective tests which the progressives had introduced. They have never reinstated objective testing.

In 1934 they adopted a plan for a rigorous academic program, but not, as in pre-Revolution days, just for the elite. Now it was to be the course of study which everybody would follow. When I say "everybody," I mean on the order of 99 per cent, excluding approximately one per cent of the population that may be mentally defective for physiological reasons. It is the Russians' belief that all others, handled properly and given the proper training, can be educated, even in such a rigorous academic program.

74

When the Russians adopted this new program in 1934, universal education extended only through the fourth grade. By 1950, they had universal and compulsory education throughout the Soviet Union through the seventh grade. Since the seventh grade in Russia is about comparable to the ninth grade in this country, this meant, essentially, education through what we call junior high school. They had planned, before Khrushchev's recent announcement, to provide universal education through the tenth grade in all republics of the Soviet Union by 1960.

Disregarding the recent announcement for the moment, let us consider the nature of the Russians' educational program today. It has, I would say, two major objectives. First, there is the goal of providing general education of the academic type for all students through the tenth grade, to supply the Soviet Union with a vast reservoir of people who are capable of further education along any line that may be necessary at a later time.

The second objective is to provide vocational training, training for specific jobs. This may be training for a semi-skilled job in a factory, where there is a fairly short course, or it may be training of a research physicist, which involves university and postgraduate work and a very long course. The Russians have a tremendous number of different kinds of educational programs and educational institutions, geared to training people for various occupations and usually much more specifically oriented toward a particular job or career than would be true in this country.

Let me describe, very quickly, this Soviet school system. It starts with the nursery schools, where children from six months to three years go while their mothers work. Next come kindergartens for children from three to six years of age. Neither the nursery schools nor the kindergartens are universal throughout the Soviet Union yet, but they are expanding very rapidly.

After the kindergarten comes the Ten-Year School, for children from seven to 17. In about half of the Soviet Union at present these schools only go through the seventh grade, after which students must continue their education by going to an evening school, a Technicum, a labor reserve school, or by correspondence courses.

In the remaining Ten-Year Schools, which do go through the tenth grade, graduating students go on to higher education or to some kind of vocational education. There are two kinds of institutions of higher education, the universities and the institutes. The universities offer work in such academic subjects as physics, chemistry, history, linguistics, and so on. The institutes are the professional training institutions, except in the case of law, which for some reason happens to be associated with the universities.

Students who do not go on to higher education may go to the specialized vocational schools for a short course, or to a Technicum for a 2½-year course of technical training. It is interesting to note that the Russians have far more of these than we do in this country. We have, it has been estimated, only about 75 technical institutes in the United States, and there are something like 4,000 in the Soviet Union, with over 2,000,000 students enrolled. Obviously, the Technicum is a major part of their educational system. As I indicated before, not all students wait until the end of the tenth grade before entering the Technicum. Some enter after the seventh grade, but they take an academic program comparable to that offered in the eighth to tenth grades of the Ten-Year School along with their technical training.

There are some extracurricular activities in Russia that are extremely important in the education of their abler students, and it would give an incomplete picture if I didn't mention them. Soviet students have clubs which they call "circles." Sometimes these are associated with the schools, sometimes with the Pioneer Palaces, which are youth centers somewhat comparable to our YMCA, Boy Scouts, and other activities all rolled into massive proportions.

At the Pioneer Palaces there are "circles" for such activities as art, music or shop work. There are also "circles" for academic subjects such as physics, mathematics, and chemistry. The leaders of these groups are usually associate professors or instructors in the universities. They encourage students interested in these special fields, and they try to spot the most able students and bring them along even faster. Often they encourage the able student to try out for the Olympiads, which are academic tournaments by subject matter fields that begin at the local level and progress through the district level to the Republic level, to the all Soviet Union level. These very highly competitive Olympiads provide a way of encouraging and developing outstanding students that is lacking in the regular Russian school program, where everybody takes the same academic courses.

Now I should like to describe in more detail the program of the Ten-Year School itself, since this is really the heart of the Soviet educational system. As I have indicated, it is an academic program and a pretty rigorous one. Perhaps I can be a little more specific if I take the last three grades of the Ten-Year School and average the number of times per week a student attends classes in a particular subject, to give you a picture of a typical year in the last three years of this school.

The student would have, on the average, six classes a week in mathematics, four in physics, three in chemistry, one in biology, five in literature, four in history (principally Soviet history), three in foreign languages (preceded by three years of study in the same foreign language, making a total of six years), two in geography, one in technical drawing, and two in physical education.

This, clearly, is a rather stiff academic program, yet the Russians expect 99 per cent of their students to take it, profit from it, and complete it. The question is: to what extent is this true? To what extent are they able to get 99 per cent of their students through a program of this nature?

I was unable to get complete statistics to provide a definite answer to this question, but trying to make the best estimate I could on the basis of the evidence available, I figured that somewhere between 50 and 80 per cent of Russian students actually get through this Ten-Year School program. In addition, others take a somewhat similar but perhaps slightly watered-down version of it in Technicums or labor reserve schools or by correspondence courses. This is something that is hard for us to understand, because we have generally thought that only 15, 20, or 25 per cent of our students could take such a program.

How do the Russians carry their students through these strictly academic subjects? A number of factors might be related to the effectiveness of their program, and I shall review some of them quickly, some in more detail.

Certainly the school buildings in Russia are no asset to the educational system. Their buildings, as everybody has reported, are drab, ordinary buildings with nothing fancy or modern about them. The same is true of school equipment such as chairs and desks, which are equally old-fashioned. But when it comes to laboratory equipment, movie projectors and screens, and slide projectors, the story is different. These are widely available and, on the whole, although I am not a terribly good judge, the equipment for laboratory experiments and the machine shop equipment seemed to me to be very good, and a real asset to the educational system.

Perhaps the strongest asset, however, and a key factor in the success of the Soviet educational program, is the Soviet teacher. Teaching is a very attractive profession in Russia. There are four or five applicants for every position in a teacher's college. The teacher's college, which is called the Pedagogical Institute, has a five-year program with thorough training in both the subject matter the individual is to teach and in methods and principles of teaching as well as in educational psychology.

Throughout their careers, teachers are expected to spend one summer out of three in further study and training. They get one day off out of the six-day school week, and are supposed to devote this time to their professional development. They are also supposed to do some outside reading in their field and report on it from time to time. In addition, there are teachers' clubs in many cities where teachers may go to consult specialists about pedagogical problems.

The normal teaching load, somewhat lighter than it is in this country, is 18 hours a week for the full-time teacher. Some do teach more than that, but they are paid extra for it. Teachers, however, do a considerable amount of special tutoring beyond their classroom work, because they are held responsible for the success or failure of their students and therefore put a lot of effort into individual work with students who are not doing well.

Textbooks and teaching aids are another factor in the effectiveness of the academic program in Russian schools. These are developed in an institution for which we have no parallel in this country—the Academy of Pedagogical Sciences. The Academy of Pedagogical Sciences, which works through eight research institutes and has a total of 550 research workers, plays a very important part in the educational program in the Soviet Union. It has access to the leading scholars and scientists throughout the Soviet Union, and can call on them to cooperate with the Academy staff on all sorts of educational problems.

One of the eight institutes in the Academy of Pedagogical Sciences is the Institute of Methods. This is the Institute that coordinates the preparation and development of new textbooks and other teaching materials. When a textbook needs revision, it is sent out to many teachers and to the leading scholars in that particular field, who are asked to comment on the textbook. Then the Institute asks a scholar or scientist who is usually one of the top people in his field to work with the Institute staff on the writing of the new textbook.

After they have worked on the problem for awhile, they put out a 100-page summary of the new book's contents, and this is also distributed to teachers and leading scholars for their reactions. After studying these, they begin to write the textbook, working on 50-page sections at a time. Each 50-page section then goes out to a substantial number of experimental schools, for trial and evaluation of how well the book gets specific ideas across to the students. Again, revisions are made, and finally the textbook is completed.

Then it has to go before an independent commission appointed by the Ministry of Education to determine whether it is better than any other text that any other individual in the Soviet Union may have written in the same field, and whether, in fact, it should be adopted for general use.

These books, if they are in the field of science and mathematics, tend to be used throughout the Soviet Union. If they are in history and literature, they aren't necessarily adopted in all the republics, because some of the republics want to emphasize their own literature and history, and they have the privilege of substituting their own textbooks.

Sound films and other visual aids are also prepared by the Institute of Methods, and are designed to supplement the textbooks in getting the most difficult ideas and concepts across to the students. The Institute has developed something like 80 films in physics, 30 in chemistry, and about 100 in biology. These films are an integral part of the respective courses and are shown in all schools. In addition, a school will have all kinds of charts. I have never seen so many charts. You can go into any classroom and there are charts on the walls depicting important points the teachers want to get across, with several hundred more charts stored in nearby closets. There are also three dimensional models to help visualize difficult concepts, and monographs available for students who want to read further in a subject.

This multitude of teaching aids is all a result of the work of the Institute of Methods. It seems to me that in preparing these materials the Russians have taken a lot more trouble than we have to make sure that school textbooks and teaching aids embody the best and latest thinking of the leaders in the field at the university level, along with the best teaching methods that those on the front line of teaching and in the Institute of Methods can devise to communicate these ideas to students.

One more reason why the Russians are able to carry a large proportion of young people through an academic program—and I think they get too many through, as I will explain later—is the motivation of the students. Their motivation is extremely high. Success in education is very important to the Soviet student. He can pretty well predict what his income and prestige will be 20 years later by the grades he is getting in school. Salaries in the Soviet Union are usually fixed in accordance with how much education is required for the job.

College professors, incidentally, are among the highest paid people in the Soviet Union. The only others who are paid more are the top party officials and government leaders. Teachers in elementary and secondary schools are also well paid. There is some difference of opinion as to how well, but on repeated occasions we were told that their salaries compared with the salaries of engineers and doctors. In any case, they must be reasonably good, or there would not be so many applicants for every opening in the Pedagogical Institutes.

So much for the factors influencing the effectiveness of the Soviet educational system. The question now is: What has been the result of this forced draught education, this massive effort to give the entire population an academic program?

What has happened is that in recent years the Russians have been training many more students in a college preparatory program than can possibly be admitted to a university or an institute. Although they have been graduating about 1,500,000 students each year from the Ten-Year Schools, the institutions of higher education can accommodate only about 250,000 full-time students and 200,000 part-time students. Thus only about one in three graduates of the Ten-Year Schools is going to be able to go into higher education. Yet all of them have been inspired with this as a goal, and they have slaved through work that for many of them was terribly difficult in order to attain that goal.

Obviously, the result is grave disappointment among those who cannot get into a university or an institute. They have to go to work, where they are needed, but they are disgruntled as workers, and they are not really prepared to do the kind of work that is required of them on farms or in industry.

When this began to become evident a number of years ago, the Russians introduced into the Ten-Year School program, on an experimental basis, what they called polytechnical instruction. Throughout the elementary and secondary grades, this instruction is given in addition to the regular academic program. Courses in handicrafts, woodworking and metal working are given in the early grades. In the later grades, there is machine shop, work with electrical machines, and periods of actual work in industry or agriculture, with students going perhaps two days a week for a certain period of time to work in a factory or on a farm. This polytechnical instruction was supposed to introduce students to actual factory or farm work, or to enable those who would go on to higher education to know how the other half lives.

But the basic problem still remained: many students were unhappy because their hopes for going on to higher education were not fulfilled; and the economy desperately needed more workers. So in April 1958, the Russians tried another idea. Khrushchev announced that thenceforth all students finishing the Ten-Year Schools should work for two years before going to the university or institute. This would get more people into the labor market and would give all students, even those who would eventually enter higher education, a chance to know what factory or farm work is like.

It was never really intended that this policy would be put into effect

100 per cent, because the Russians knew that they had to let their mathematicians and scientists continue their studies without interruption. The fact of the matter is that 50 per cent of the students admitted to the University of Moscow and the University of Leningrad last year came directly from school. But the announced policy made the students who had to go to work feel that this *was* policy, and that they were doing something that was both good for them and good for the country.

Since last April, however, the Russians have begun to realize that neither the polytechnical program nor requiring students to work two years before going on to higher education was going to solve the grave problem that lay ahead of them. The Soviet Union is up against a situation that is completely different from the one we face. We are coping with a big bulge in the birth rate that began during World War II, and is just now filling our high schools and colleges with more students than ever before.

Russia, on the other hand, suffered a decline in the birth rate during the war years, and therefore has fewer rather than more young people coming through its educational system and entering the economy at the present time. The figures on this are quite dramatic. A year ago, there were 6,250,000 17-year-olds in the Soviet Union. Next year, just two years later, there will be only 3,250,000 Soviet youth in the 17-year-old age bracket. Of the 6,250,000 17-year-olds a year ago, 1,500,000 were regular students in the tenth grade of the Ten-Year Schools, leaving 4,750,000 available for work in factories and on farms. But next year, when there will be only 3,250,000 17-year-olds, and 1,500,000 of these will be in school, there will remain only 1,750,000 available as workers —3,000,000 less than they had a year ago.

In view of the rapid rate at which the Soviet economy is expanding, it is clear that the Russians must have more workers available immediately and cannot afford to allow so many students to continue with their education at this particular time. This constitutes a grave problem, and, I suspect, is one of the factors behind Khrushchev's latest announcement. A little more than a month ago, in September 1958, Khrushchev issued some new proposals which, he said, were set forth by the Praesidium of the Central Committee of the Communist Party. He also suggested that these proposals be discussed up and down the land, so that reactions could be considered before final plans were adopted.

The basic idea of the proposals boiled down to this pronouncement: "All boys and girls without exception will go to work after the eighth grade, and do their studying on a shift basis or in the evening by correspondence." If this were actually put into effect, the shortage of

81

young people available for work would obviously be somewhat relieved, because there would be a couple of extra age groups quickly poured into the labor market.

But there are two interesting points worth noting in this latest Khrushchev announcement. First, he contradicted within the same announcement the statement "without exception," and said that the very able students in the arts and music, in mathematics and the sciences, would continue to study full time at special schools.

Second, he stated that those who go to work after the eighth grade and later apply for admission to a university would be admitted on the basis not only of examinations, but also of recommendations from the trade union and the Young Communist League.

Khrushchev noted that 70 per cent of those going on to universities now had come from families of intellectual workers and office workers, and only 30 per cent from the vast population of industrial workers and farmers. He declared that these figures must indicate something wrong in the admissions system of the universities, and that the situation should be corrected. What he did not say, but probably meant, was that he and other Communist leaders are concerned about the possible development of a class society, of an intellectual elite resulting from the fact that intellectuals are having children who become intellectuals, and that, sooner or later, these generations of intellectuals might threaten the Communist Party leadership.

It seems likely, at any rate, that two new hurdles will be introduced for a large proportion of Russian students applying for admission to institutions of higher education: they will have to get the approval of the trade union, which means they must have worked hard and been good workers, liked by their fellows; and they will have to win approval from the Young Communist League, which must feel that they are either good potential Party members or at least the kind of people in whom the Party can have confidence.

Now, I doubt that these developments indicate any less devotion on the part of the Russian people, or the Soviet leaders, to education. What they do indicate is that Russia has to contend with some grave problems, particularly the shortage of workers, and that it must make some adjustments in its educational system, at least for awhile. These developments may also indicate that the Russians' overenthusiasm for general education of an academic type for everybody, regardless of ability, is going to be tempered.

Members of this conference will be interested, I think, in two topics I have not yet touched on: *guidance* and *examinations.*

Guidance, as it is known in this country, is nonexistent in Russia, and in fact the term is not understood. They do not use tests and they do not have guidance counselors. There is less necessity for either since every student takes exactly the same rigidly academic program. They do have, however, outside of the regular school program, a system which might be called "guidance by exposure." This is the tremendous variety of opportunity for extra-curricular activities offered by the "circles," or clubs, in the Pioneer Palaces. Students have ample opportunity to try out possible interests by joining one or another of these clubs. Since the clubs start at the first grade level and continue to tenth grade, there are opportunities for each student to find an activity of interest to him and in which he can do well. This, so far as I could find, was the only method that the Russians have for guiding students into special fields.

Examinations in Russia are greatly different from those used in the United States, and seem very strange to those of us familiar with the kinds of examinations that are used in this country. Objective tests cannot be found in Russia, except in a few psychological laboratories. They were used in the early years of the Communist regime, as I mentioned earlier, but were thrown out in the early thirties along with the progressives who had used them. The principal kind of test used in this earlier period was the IQ. Since, by its very name, an Intelligence Test implies the great importance of heredity, as opposed to training, the IQ test contradicted Party doctrine. In the Soviet Union, it is the firm conviction of the "powers that be" that background and training, particularly the latter, are far more important than heredity. They recognize that individual differences do exist, but claim that all individuals can be brought to a remarkable level of achievement by proper training.

IQ tests not only contradicted the accepted doctrine but, even worse, operated to select the children of intellectuals rather than the children of the favored "workers." When this continued to happen after a generation of the Communist regime, it meant either that the Communists had not provided the kind of society and education for the "workers" that they had set out to provide, or that their theory was wrong and inherited ability was, in fact, of major importance. Neither conclusion was acceptable and all tests were banished.

If the Russians don't use objective tests, what kind of examinations do they use and what do they use them for? Tests are not used for identification, guidance or placement of individuals as they are in this country. The main use of examinations in Russia is to audit achievement. Principal reliance is placed on oral examinations of a rather special type and to some extent on written examinations. The purpose of these exams seems to be limited to assuring that the essential elements of the important courses have been completely mastered.

Examinations prepared by the Ministry of Education used to be given at the end of each grade, but gradually this program has been reduced until at the present time examinations are given only at the end of the seventh grade and at the end of the tenth grade. There are also examinations for admission to the universities and the institutes. They are quite similar to, though more demanding than, the examinations given at the end of the tenth grade.

At the end of the seventh grade, all students take a written examination and an oral examination in Russian and a written examination in algebra. At the end of the tenth grade, there are written examinations in Russian literature and composition, and oral examination in mathematics, physics, chemistry and Soviet history. The unusual feature of the Russian examinations can perhaps best be explained by describing one oral and one written examination. For the tenth grade oral examination in solid geometry, for example, a class of 30 is divided into two sections of 15 each. All 15 students go into a classroom where the teacher of the class, the director of the school, one or two other teachers, and sometimes representatives of the educational authorities of the city sit as an examining board.

On the table, at which the board sits, are 20 to 25 "tickets," each of which contains three questions. Two of them are rather standard proofs or problems and the third a problem that, while not new in type, is perhaps a little different from the problems that have been assigned during the year. All three questions are concerned with a particular topic, and each of the 20 to 25 "tickets" covers a different topic. Students therefore have to be sure that they have covered all the topics, since they can never tell in advance which questions they may be called upon to answer. Fortunately, for them, the Ministry of Education publishes a pamphlet several months before the examinations in which the topics to be covered are listed, and in some instances the first two questions are specifically stated. Only the third problem or question remains in doubt.

At the beginning of the examination four or five students select their "tickets." Then they return to their desks and work out the answers. When the first student is prepared, he goes to the blackboard and writes out his answers on the blackboard. He then explains his answers to the examiners and responds to any questions they may have. The examiners may ask questions about any aspect of the course, but

I gather that this privilege is not exercised to any great extent. Meanwhile, of course, students two, three, and four have had considerable time to prepare their answers, but at the same time they have had a good deal of distraction. Students 13, 14, and 15 have to sit through a long morning of waiting until their turns come up.

It is evident from their examination system that the Russians focus attention on the mastery of important topics to a much greater extent than we do in this country. On the other hand, there is far less incentive to study more than the minimum essentials. The examinations are primarily a motivating device. They provide a goal toward which the student strives. In the schools we visited, it was rare to find that any student had failed his examinations the previous year. Teachers may prevent students from taking the examinations if they have had a low average during the year, but as far as we could ascertain only one or two students in a class of 30 would be prevented from taking the examinations.

On written examinations all students are presented with the same questions. Each student must answer only one out of three questions. The topics tend to be very general. Students write an essay of not more than about eight blue book pages. Surprisingly enough, they have six hours in which to do this. The examination in Russian literature, which I observed, began at 9 o'clock in the morning and continued until 3 p. m. It was customary, and perhaps even mandatory for each student to write a first draft and then, having worked it over very carefully, copy it before the end of the examination. Our own experience with essay examinations would lead us to question the time limits set for the examination, and the reliability of an examination involving only a single question. However, the Russians' objective seems merely to determine whether students have met a certain minimum standard of literary interpretation and writing.

The growing interest in cultural exchange and comparative education may some day make possible the administration of suitable tests, both in this country and in Russia, to groups whose preparation has been comparable. It would be particularly illuminating to see if there are significant differences in the performance of American and Russian students, and, if so, how they differ. On the basis of present knowledge, one might hazard the guess that in cumulative subjects such as mathematics, foreign language, and the sciences the Russian students will have a better command of minimum essentials, whereas American students will have a broader view of the subject and perhaps somewhat more flexibility in applying the knowledge they have gained in a variety of situations.

HENRY CHAUNCEY

Much of the education in the Soviet Union at the present time seems peculiar to us. Many aspects we would criticize. Nevertheless, we have to recognize that here is a country that has made tremendous strides in education during the past forty years. The people have had to overcome great obstacles. They have been tackling their problems with imagination and they are experimentally-minded. They are flexible and adaptable; they are energetic in putting changes into effect on a nationwide basis; and most important of all, they have a tremendous belief in education. They will press forward in the educational field—not to make life full for the individual, not to bring about his full development so that he will lead a rich and rewarding life, but to make a strong, powerful country to reach and surpass America scientifically, militarily, and economically.

Session III

Remarks of the Chairman

We come now to the final session of this Conference, and something of a change of pace as well as a change in topic. This session will be a panel discussion on one subject with four different speakers presenting diverse views on the problem posed.

We have all been hearing, until the expression has become a cliche, of the impending tidal wave of college students. What the tide is going to be, high tide or low tide, is going to depend in considerable part upon the selection and admissions procedures that are adopted by our institutions of higher learning.

The determination of the optimum types of examinations for college admission and scholarship purposes is certainly one of the most pressing problems in the measurement field. We have brought together this afternoon, for a thorough airing of the issues involved in this problem, four persons who have devoted long and careful thought to these matters, and who have engaged in extensive research on the topic. I am confident they will cover the field with comprehensiveness and that they will bring to your attention the pros and cons of the varying points of view on the testing needs in this area.

It is a pleasure to introduce to you our four panel members: Dr. Robert L. Ebel, Vice President of the Educational Testing Service, will lead off with his views on this afternoon's big question, "What Kinds of Tests for College Admission and Scholarship Programs?" The second panel speaker is Dr. John C. Flanagan, of the American Institute for Research and the University of Pittsburgh, who will discuss "Criteria for Selecting Tests for College Admissions and Scholarship Programs."

Next will come Dr. E. F. Lindquist, of the State University of Iowa, who will focus his remarks on "The Nature of the Problem of Improving Scholarship and College Entrance Examinations." And our final speaker is Dr. Alexander G. Wesman, of The Psychological Corporation, who will summarize for us his answer to the question, "What Kinds of Tests for College Admission and Scholarship Programs?"

What Kinds of Tests for College Admission and Scholarship Programs?

Robert L. Ebel, Vice President, Educational Testing Service

The use of tests for college admission and scholarship programs has grown rapidly in recent years, and seems likely to continue to grow in the years immediately ahead. If so, the wisdom and experience of you who are here today will have an important bearing on the future effectiveness of American higher education. I am grateful for the opportunity of taking part in the discussion of one problem associated with this development, and am honored to be included in so distinguished a panel. You will understand, I am sure, that on an occasion such as this, I will try to express my own views as frankly and clearly as possible, and that I will not presume to speak for ETS or any other institution or group.

What kinds of tests should be used for college admission and scholarship programs? The answer to this question is, in principle, quite simple. We should use the tests which are the most valid. Some people think the answer, in practice, is equally simple. Choose the test or combination of tests which gives the best prediction of first year college grades. I demur. It seems to me that this approach is not likely to yield adequate evidence concerning the relative merits of different kinds of tests.

There are at least three limitations of conventional validity studies in this situation, as I see it. First, the criterion of college grades or grade point averages is itself unreliable and of quite imperfect validity. Second, sampling errors associated with the obtained validity coefficients tend to be so large, and precise estimates of those sampling errors so difficult to achieve, that it is almost impossible to make dependable comparisons of the merits of alternative tests. Third, and most important, this approach assumes that a college's current educational program is beyond criticism or improvement. If students who score high on the selection test do poorly in college, the selection test rather than the college program is blamed. I submit that there are better ways of improving the input to our colleges than by striving to improve the prediction of faulty measures of student success in attaining poorly defined and somewhat questionable goals.

One of these was described by Ralph Tyler, writing on "Educability and the Schools" in the Centennial volume of the American Association for the Advancement of Science. He contrasted the prediction-of-success approach with a development-of-talent approach that would seek to capitalize on the important latent abilities revealed by appropriate tests. He suggested that our present school programs do not capitalize on all such abilities, and that tests designed only to predict success in current programs of instruction do not adequately measure the characteristics which determine educability. I heartily agree.

There are some who mistrust subjective decisions concerning the nature of these characteristics, preferring tests which can be defended on the basis of their high correlation with some ultimate and presumably more objectively given criterion. But the only truly ultimate criterion of success, if indeed there is any such thing, lies hidden beyond the grave. And the apparent objectivity of some criteria hides the subjectivity of our decision to use them as criteria. What reason have we to believe that our subjective decision to use more immediate criteria of success, such as wealth, or grades in the freshman year of college, are any more trustworthy than our subjective judgments concerning the elements of a good preparation for higher education?

Even if we are willing to overlook the possible imperfections of our criteria and experimental designs, why should we employ an approach which allows the inevitable contingencies affecting success—health, finances, motivation, even romance—to blur and becloud the application of these judgments to the choice of admission or scholarship tests. Would it not be better to apply the inescapable acts of judgment to the tests directly? And is this not actually what we do? Were not most of the college admission and scholarship tests in current use designed and built on the basis of rational inductions, deductions and hypotheses? Empirical procedures surely were used to defend them, and to refine them, but the basic structure seems usually to have been determined, as I think it should have been, by purposeful cogitation rather than by completely objective, judgment-free experimentation. Hence, what I have said about the limitations of the grade criterion and the conventional validity study should not be construed as a complete rejection of such studies. Some of our judgments about what we want to measure, and what we have succeeded in measuring, can be checked empirically. But I am convinced that validity studies should not be the exclusive, or even the primary, basis for test selection. One test or battery of tests should not be chosen over another simply because of a small apparent advantage in predictive validity.

Among the vast array of tests which might be used in admission and scholarship programs three major types can be identified: tests of innate capacity for learning, tests of developed ability, and tests of substantive knowledge.

Many people regard tests of innate capacity for learning as nearly ideal for college admission and scholarship programs. The term innate mental capacity suggests something which is fundamental and permanent, and hence well worth measuring, a divinity that shapes our ends, no matter how poverty, inadequate schools, or youthful follies may have rough-hewed them. A test of innate capacity presumably will not handicap the bright youth who grew up on the wrong side of the tracks. In theory, scores on it should not be affected appreciably by coaching, or indeed by instruction of any sort. Such a test would place no restrictions upon the secondary school curriculum. Local schools and individual teachers could presumably retain their freedom to teach what they choose, and it would make no difference if they taught it well or badly.

Unfortunately for this particular vision of utopia, no one has found any very accurate way to measure innate capacity for learning. All of the alleged measures of this capacity are more or less obvious measures of educational achievement. The main differences among them arise from the varying degrees of earnestness with which their authors have attempted to avoid measuring the results of school learning. The kinds of tasks a student is asked to perform in taking an intelligence test are tasks which he has been taught to perform in school, or could be taught there if the school considered them of sufficient importance. Lacking precise control of educational influences, pre-school, in-school, and out-of-school, we cannot tell how much of a student's success on an intelligence test should be attributed to his native mental capacity, and how much to his subsequent learning. To regard scores on such tests as acceptable measures of innate capacity for learning requires assumptions which I find hard to accept.

Even if tests of innate capacity for learning were available, they probably would not be desirable for college admission and scholarship programs. Innate capacity is not directly relevant to ability to profit from college instruction. Unless this innate capacity has been developed —unless the student has acquired considerable knowledge and numerous abilities—he is unprepared for college no matter how large his empty innate capacity may have been.

Some fear that unless we use intelligence tests we may miss bright students who, growing up in an educationally deprived environment, show little achievement and hence might be denied an opportunity for college education. Evidence which would justify this fear is hard to find. Many cases of brilliant adults who were unpromising youthful scholars can be cited. But efforts to identify such individuals in advance by using tests of mental capacity have been disappointing. Errors of measurement can account for much of the observed discrepancy between measures of so-called ability and achievement. Differences in the kind of achievement required by the two types of tests can account for much more. Aware of these facts, most of us have ceased to trust achievement quotients, but some perpetuate the same fallacies searching for over- and under-achievers. Some "under-achievers" go on to college success. So do some over-achievers. Intelligence tests do not help very much in identifying brilliant minds which the schools have missed, or failed to develop. All learning builds on previous learning. If the foundation is weak the superstructure is likely to suffer. What the colleges require are students who have strong education foundations, not those possessing brilliant but undeveloped minds.

One alternative to measurement of native mental capacity as a basis for college admission and scholarship awards is the measurement of a student's command of essential knowledge. This alternative has not been popular in recent years. When knowledge is contrasted with ignorance, it is universally praised. But knowledge is sometimes contrasted with understanding, with wisdom, or with character, and in these contrasts knowledge fares badly. Teachers speak disparagingly of "stuffing the mind with facts." College presidents emphasize the limitations of "mere knowledge," and stress the contributions that a college education can make to the development of character. Quiz kids, or even adults who show an unusual scope and accuracy in their recall of isolated items of information, are used to illustrate what a good education does *not* consist of. "Scraps of information have nothing to do with culture," Whitehead has said and continued, "A merely well-informed man is the most useless bore on God's earth."

Attacks like these on pedantic, trivial, verbalistic, unassimilated knowledge have been at least partly responsible for general reluctance to use tests of substantive knowledge in college admission testing programs. But I would suggest that this policy deserves reexamination. Quite obviously, knowledge can be defined so narrowly, or caricatured so grotesquely, that all of the above attacks on it will seem to be well founded. But knowledge need not be limited to isolated, trivial, informational details, nor to verbal abstractions divorced from reality, nor to rote responses to stereotyped questions. It can be defined broadly enough to include understanding, wisdom, and other approbatory synonyms for effective rational behavior. Understanding, for example, consists in *knowing* the interrelations between other items of knowledge. Wisdom consists of *knowing* how to use the knowledge one possesses. Character, insofar as it is expressed in rational rather than blindly imitative or authoritatively imposed behavior, is based partly on *knowledge* of the consequences of alternative courses of action. It is in this broader context that one can truly say that knowledge is power, and argue that it represents the primary outcome of the educational process. Quoting Whitehead again, "Education is the acquisition of the art of the utilization of knowledge." I suggest that the "art" Whitehead speaks of is itself essentially *knowledge* of how best to proceed in a given set of circumstances.

Apart from unwarranted restrictions on the concept of knowledge, there is another reason why knowledge tests have not been widely used as a basis for college admission testing. The scope of human knowledge is so broad, and the areas with which different individuals are acquainted are so diverse that it seems difficult to construct any single test, or even any limited battery of tests, which can deal adequately with this abundance and diversity. Our decentralized school system, and pupil-centered teaching procedures tend to foster wide differences in the kind and level of education that different pupils receive. But since we live in the same society at the same period in history, all of us need to know and to be able to do many of the same things. It may be difficult but it is not impossible to define a common core of essential knowledge.

There is a third alternative to tests of mental capacity, and tests of command of essential knowledge, for college admission and scholarship testing. These are tests of mental traits or developed abilities. Proponents of tests of this type refer to them as measures of broad intellectual skills, of basic mental processes, or of habits of thinking. It is said, in their behalf, that they measure what a student is able to do, more or less independently of what he knows. The abilities they purport to measure range from the highly general, such as "ability to think" to the fairly specific, such as "ability to formulate hypotheses."

Insofar as these tests emphasize essential rather than trivial achievement, require the student to show understanding rather than mere recall, and ability to use rather than mere ability to reproduce, they deserve enthusiastic applause. But when they are represented as measures of distinct mental processes or independent intellectual skills, something different from and more important than knowledge, which can be developed by practice with a wide variety of content materials, they seem to call for critical examination.

One thing about them which is troublesome is the vagueness of what they refer to. There is no lack of names for alleged mental abilities, but there is almost universal lack of specific definition of what these names mean. The evidence that these abilities are independent of knowledge, representing a kind of mental function which can be applied at will to knowledge in diverse areas, is practically non-existent. They suggest a renaissance of the once discredited belief in faculty psychology which held, for example, that memory for something like faces could be improved by practice in remembering something else like spelling words.

If proponents of the developed abilities approach to mental measurement regard such things as ability to calculate a square root, or to diagram a sentence, or to locate a malfunctioning element in a television circuit, as developed abilities, I have no quarrel with them. We are simply using different words for the same thing when they call them developed abilities and I call them command of substantive knowledge. But when proponents of tests of developed abilities speak of highly generalized, undefined or vaguely defined complex higher mental processes or mental traits, we part company. By using high-sounding terms we may impress outsiders but we are not likely to contribute much to the improvement of mental measurements.

Many of the so-called "hard-to-measure" qualities appear to fall in this category of developed abilities. Few would deny that terms like "creativity," "flexibility," "sensitivity," or "balanced judgment" are useful in describing behavior. But we make a tremendous logical leap when we assume that they are also names for mental traits which help to determine the behavior observed. Before agreeing that our failure to measure these "traits" is a blemish on the record of mental measurements, I would like to be a little more certain that something exists to be measured. Skinner, Holland and others have pointed out that psychologists have a weakness for inventing explanatory concepts to account for meager or non-existent data. Perhaps this tendency, which many educators also share, is evidence for a hard-to-measure mental trait called "hallucinatory imaginativity."

The developed abilities approach to the measurement of educational aptitude is based on an allegorically attractive, but experimentally unsubstantiated conception of how the mind functions. All that we have learned concerning the process of learning in rats, monkeys, and men, can be explained in terms of the formation and destruction of associations among perceptions, concepts and ideas. Electronic brains, including primitive models capable of learning or even of self-reproduction, operate essentially on the basis of switches which open or close alternative circuits. To regard the mind as a muscle which is strengthened by exercise, or as a processing organ which can be trained to see relationships, to solve problems, to analyze, to think critically, or to create, irrespective of what kind of problems are given to be solved, what materials to analyze or think critically about, or what product to be created, simply does not accord with what we now know about brain functions. What we do know strongly suggests that educational development consists essentially in the accumulation, integration, and ready-reference-indexing, of knowledge. By practice in using knowledge, one acquires command of it. This, it seems to me, is the essence of educational achievement.

Tests of developed ability are sometimes offered as a method for keeping peace between test constructors on the one hand, and curriculum builders, school administrators and teachers on the other. For it is suggested that tests of this type measure how well a student has been educated without regard for anything specific that he may or may not have learned. This, it seems to me, involves a contradiction, since the quality of a person's education cannot be independent of what he has learned. Any test which discriminates between well and poorly educated students will inevitably reward the curriculum builders, school officials and teachers who have been most successful in imparting a good education. Of course a test which rewards those students who have learned some arbitrary selection of unimportant details is not a good test of quality of education. To qualify for this designation, it must measure the degree to which a student has achieved command of the most important general ideas and skills. It is not easy to reach agreement on what the most important ideas and skills are, but we cannot make good tests if we try to dodge the problem. And we should not mislead test users into thinking that we can.

Test builders obviously should not dominate the curriculum or dictate what teachers should teach. But someone must define common educational goals specifically enough to permit determination of the extent to which they are being reached. In this endeavor the test constructors can be essential allies of curriculum builders, school administrators and teachers, especially if they concentrate on measuring a student's command of essential knowledge.

It is quite apparent that tests of developed abilities do not actually succeed in measuring mental processes apart from the examinee's knowledge or lack of knowledge of particular items of information. They do measure knowledge, but since they are designed to measure something else, they often provide poorly balanced, inefficient tests of knowledge. The assumed necessity of supplying essential background information in the test itself leads to relatively verbose tests, which give undue weight to reading skills, and require undue amounts of time per scorable response. Philip Vernon suggests that tests of this type may yield a test form factor which seriously biases the scores from such tests. The score a student obtains may depend considerably on his knowledge of how to handle the particular type of task these tests present. Further, the use of background materials leads to tests composed of clusters of items, which restricts the freedom of sampling in the test and complicates the processes of item selection. Hence I am persuaded that a well-designed test of substantive knowledge which measures a student's command of basic useful knowledge will provide more relevant information concerning a student's educational achievement and aptitude, and provide it more efficiently, than current tests of educational development or developed abilities.

Although these tests of complex mental processes and developed general abilities have been in use for more than a decade, there is little evidence that they measure something more important than can be measured by tests of a student's command of substantive knowledge. Some of their advocates argue that such evidence is unobtainable in principle, and that the only way to ascertain their superiority is by looking at the tests themselves. But the main job of a test is to yield useful scores from examinees. If the scores from one test are better than those from another, they must at least be reliably different. If the correlation of scores on a superior and an inferior test is as high as that between two forms of the superior test, then it seems to me that all of the claimed superiority is getting lost in the errors of measurement. I do not see how such a test can be said to be superior functionally as a measuring instrument.

A desirable characteristic of a college admission testing program is that it have a stimulating, constructive influence on the programs of instruction in preparatory schools. Obviously a test of mental capacity could not have this influence. Tests of developed ability because of the generality and indefiniteness of what they are testing, are unlikely to stimulate either teachers or students to unusual efforts to improve. Since the first Sputnik orbited in space, the practices of the Russian schools have received considerable attention. It would be foolish to suggest that we should copy exactly what they are doing. But it would be equally foolish for us to ignore completely their accomplishments, and the means by which these were achieved. Henry Chauncey and others who have visited Russian schools report with some admiration the persistent and generally effective efforts of teachers and students to achieve limited but clearly defined goals in achieving command of substantive knowledge. If we wish our college admission and scholarship award programs to have the most beneficial effect in the improvement of public education, it would seem wise to emphasize tests in which the nature, and above all the quality, of the teaching done by each teacher makes a direct and obvious difference in the scores the students receive.

I suspect that my colleagues on this panel do not share completely my admiration for tests of substantive knowledge. If it were possible to do so, I would like to test our divergent views with a little experiment in which we would compete with each other in selecting capable students.

Suppose that one of us prefers a test of mental capacity, another some tests of developed ability, and yet another a battery of pure factor tests. My preference, of course, would be for a test of relevant substantive knowledge. Suppose that we have available to us a population of 100 pupils who are just ready to be taught some new process, such as the solution of simultaneous linear equations in algebra. Assume that each of us is a tolerable teacher of the process the students are about to learn. We each give our preferred aptitude test to all of the students, and then proceed to choose up classes. After choosing our classes, and after an agreed-upon number of periods of instruction, we would give our students a common test of achievement and determine which of us had done the best job of selecting students, and of teaching them.

This is a hypothetical experiment, not only because we are unlikely to have an opportunity to perform it, but also because I seriously doubt that any of the other panel members would wish to rely, when the chips are down, on a test of mental capacity, developed ability, or pure factors, to select students with the greatest aptitude for a particular job of learning. I suspect we might all agree that in this situation a specific test of substantive knowledge would be more effective in selecting students than a general test of capacity, of developed abilities, or of mental factors. My colleagues may object that I have prejudiced the argument in my favor by directing attention to a single specific problem of learning and teaching rather than on the diversity of such problems which face the student admitted to college, and his teachers. This is a reasonable objection to drawing general conclusions from the experiment I proposed, but it has interesting implications. It suggests that tests of mental capacity or developed abilities should be used, not because they measure something more basic to effective learning than acquired knowledge, but because they provide convenient *general* measures of competence. Perhaps they do, but I am inclined to doubt it. Unless I am mistaken about the phantom nature of many mental traits and developed abilities, and the substantive knowledge nature of the rest, the best way of measuring a student's preparation for college learning *in general* is to set out directly to measure the most generally useful aspects of substantive knowledge.

This afternoon I have questioned some widely held opinions about aptitude tests. If what I have said sounds dogmatic, the reason is at least partly that you and I have not yet done enough research on fundamental problems of aptitude testing to provide a more substantial basis for our beliefs. I am far from believing that the views I have expressed are the only ones any reasonable man can entertain. In fact it is now time for me to yield to another reasonable man who will present some other views for you to consider.

Criteria for Selecting Tests for College Admissions and Scholarship Programs

JOHN C. FLANAGAN, American Institute for Research and University of Pittsburgh

The basic assumption underlying this discussion is that in our country the collective aim is to enable each individual to realize his highest potential. This contrasts sharply with the Russian view that the aim of education (and all other activities) is to make the country as powerful as possible.

Since individual talents cannot be expected to develop properly in our complex society without systematic, appropriate, and extensive nurture, it is of utmost importance that the individual's talents be identified early as a basis for his educational and career plans. This point of view suggests that college admissions should be viewed as a part of an overall student counseling and guidance program.

It follows that colleges and universities and other organizations sponsoring scholarship programs might most appropriately enlist the aid of counselors, secondary school teachers, and other school officials in assisting them in determining which students should pursue their education under the auspices of each of these colleges. There are several fundamental considerations which have contributed to this conclusion. Each of these will be discussed briefly.

1. *A clear statement of the objectives of the college, university, or other organization should be the ultimate basis for college admissions and scholarship policies.* Unless the institution knows precisely what it wishes to accomplish, there can be no evaluation of its success or failure. For this purpose, a general statement of objectives will not suffice. The broad aims of the institution must be translated vividly and with detailed examples if they are to provide a practical framework for developing selection policies.

Of course, broad statements of aim are needed as a basis for developing specific aims. A good example is the general aim quoted from the remarks of the president of one of our leading universities. In dis-

98

cussing the objectives of a college administrator, he proposed that the aim should be ". . . to help every young person in his care grow into the broadest, deepest, most vital person possible. And in fulfilling himself, the student will . . . arrive at moments of heightened insight when he sees more clearly than ever before what the world is about and how he can fit into it creatively and significantly."

Such a general aim must be translated into specific activities of faculty and students if it is to be useful for formulating student selection procedures. Detailed statements about the dynamics of such student growth can contribute directly to decisions regarding selection policies and procedures.

Insofar as colleges reject the objective of "just cramming students with facts in order to teach them how to earn a living," college grades and achievement in typical college achievement tests cannot be regarded as a satisfactory criterion with which to evaluate selection procedures. The colleges must do something about developing working statements of their objectives if research answers are to be obtained.

2. *Scholarship and college admission policies might well be regarded as an integral part of a broad program of individual guidance.* It is proposed that, lacking the detailed statements of college objectives referred to above, this aim be defined as "to make it possible for each young person to identify and obtain the education necessary for him to realize his individual potentialities and to gain lasting personal satisfactions."

The primary implications of this consideration are that the guidance counselors and the college admissions officers should regard themselves as a team working together to achieve the objectives of both the individuals and the institutions to the greatest extent possible. In order for this team to function effectively, college admissions officers should communicate detailed information to the secondary school counselors regarding the specific opportunities for educational development at their institutions. This information should include reports concerning the characteristics of the students who benefit most from these opportunities.

The function of the counselor is to collect and communicate data regarding individual students of the types identified by college officials. Psychologists and educational measurement specialists should make every effort to develop satisfactory tests and related procedures to aid counselors in this task. However, in the absence of satisfactory psychometric techniques, counselors should use all informal data gathering procedures available to provide as good an estimate as possible of the students' characteristics.

3. Consistent with the first two points, *the primary criterion for evaluating selection policies and procedures is the performance of the individual after he leaves the college.* This statement does not mean to imply that performance in an occupation is to be taken as the only criterion. It is intended that performance with respect to all aspects of the life of an educated individual including citizenship, parenthood, cultural and personal development be included. This focus on life rather than on schooling for evaluating admissions policies casts further doubt on the adequacy of current procedures.

Two recent studies, for example, show little relation between predictions based on ability and achievement and subsequent performance. The first of these is Terman and Oden's follow-up study of gifted children. In this study the conspicuously successful group and the relatively less successful group showed only a slight difference in average intelligence test scores. The other study has been reported by Harmon. This consists of a follow-up of a group of applicants for scholarships. Committees of professors using college records, recommendations, and other information selected some of the group for scholarships and rejected others. It is reported that several years later, at the time of the follow-up, the average performance after graduation of the individuals in the two groups was nearly indistinguishable, according to the appraisals made by new committees. These new committees consisted of persons of the same type as had made the original appraisals. In this study, grade point averages and tests of verbal ability, quantitative ability, and educational achievement all showed predictive validities of approximately zero. It seems inappropriate to continue to rely on these general measures.

4. *The pattern of aptitudes required for success in each of the important career fields is relatively specific to that field.* Because of the dependence of educational instruction on verbal comprehension and the ability to solve quantitative problems, it has been assumed that these abilities are the primary determinants of successful performance in nearly all the career fields for which college training is provided. One of the early refutations of this point of view was the successful discrimination between the aptitude patterns required for successful performance of pilots, navigators, and other aircrew members in the Air Force during World War II. In 1941, officers responsible for selecting pilots were selecting applicants on their knowledge of history, literature, ability to read, and vocabulary. Some of these measures were found to have slight negative correlations with success in pilot training. The widely held view that a person with average ability in dealing with verbal and quantitative materials can succeed in practically anything if he merely

applies himself is gradually being replaced by the more sophisticated view that a large number of specific aptitudes play important roles in determining performance in the important career fields.

5. *Interest and motivation factors are at least equal in importance to aptitude factors in determining performance in specific career fields.* It is well established that the top performers in many fields are those most strongly motivated with respect to that activity. The effort and persistence which are observed to accompany outstanding effectiveness are based on a high level of interest. In the Army Air Force during World War II, it was found that the best single predictor of success in pilot training was a measure of interest and motivation in the form of an information test.

A specific example of the effects of insufficient interest and motivation is provided by an American pilot who was tested as an employee of a foreign air line. This pilot had been trained during World War II in the U. S. Army Air Force. After the war he went back into the department store business. However, because of the large salaries available in commercial flying, he accepted an offer to fly for the foreign air line. His aptitudes were found to be relatively high with respect to all of the abilities required of air line pilots. On the tests having to do with interest and motivation, on the other hand, his scores were unusually low. His pattern of interests was strongest in business, accounting, literature, and the fine arts. He showed almost no interest in mechanical problems in general or aviation matters in particular.

The chief pilot of the air line reported that the pilot in question spent most of his time reading poetry or classical literature while his co-pilot flew the airplane. He made no effort to learn about new equipment or devices, or to maintain his flying skills. They reported that in the two years he had been flying for their air line he had twice failed to pass his six-month instrument check. Each time he had been taken off flying and given special training. Each time he responded very quickly to the special training and was soon able to pass his check flight and go back on flying status. Clearly, such an individual is miscast as a pilot and is not only ineffective, but a potential hazard to company equipment and passenger lives. Many instances of the reverse situation in which a high degree of interest and motivation have more than made up for specific aptitude deficiencies have also been observed.

6. *A comprehensive program of research is required to identify the talents needed for various careers, to determine the effectiveness of various types of education in developing these talents, and to formulate the best procedures for assisting individuals in defining their roles and planning for the most effective and satisfying use of their talents.*

Isolated studies carried out during the past thirty years have identified certain types of behavior measures as essential for effective performance in specific jobs. However, systematic and comprehensive studies have been confined primarily to certain job areas in the military services. Even less has been done in evaluating the comparative value of various types of educational experience for effective performance in particular careers. Least of all is known about which counseling and guidance procedures are most effective in assisting students to develop a realistic self-concept and plans for attaining their goals. Although many types of research studies can be expected to contribute to knowledge regarding these matters, it is believed that most progress can be expected from a comprehensive, large scale, long-range project using electronic scoring and data processing techniques. The planning phase for such a study has been nearly completed and partial support for the main study has now been made available. As a preliminary indication of what such a study may reveal, some results on a relatively small follow-up study on high school seniors in Pittsburgh are reported here.

On the basis of a follow-up of 1016 persons five years after they were tested as seniors in high school, it was found that 329 of this group entered college. Of this group, 193 had received a bachelor's degree from a college at the time of the follow-up. About 30 percent (95) had quit college before completing their courses. The others were still enrolled and expected to complete their courses. About 60 percent of those entering college graduated or expected to graduate in the course they had entered. Only about 25 percent of those going on to college took the course and entered the occupation planned on while in high school.

Many of those dropping out or changing courses would have been advised not to enter these courses on the basis of their pattern of aptitudes alone. In a few cases their combined aptitude score for the specific course was as much as 1.5 standard deviations below the minimum recommended. Such wasted effort is all too frequent under the present system.

7. *Decisions concerning which students should attend college, and which college they should attend to attain their objectives should be arrived at by a process of successive approximations.* Although the necessary data on which to base these decisions is only partially available, colleges and students must continue to make such decisions It therefore seems most appropriate that both the college and the student begin early with a tentative set of decisions and continue to obtain as much relevant data as possible until the final decisions regarding the student, his college, and his courses must be made.

For example, it seems appropriate that as early as the 9th grade the student, with the help of the counselor, should begin to list possible career choices and possible colleges and courses which would be appropriate for him. College admissions officers could provide the counselors of 9th grade students with information regarding tentative standards, not only with respect to aptitude and achievement, but also with respect to personality, interest, motivation, and activity factors. Each year during the secondary school period, the possibilities should be reviewed and the student should set certain goals for improving his information regarding his own aptitudes, interests, motivations, and other personal characteristics as related to college requirements and college opportunities.

It is believed that such a procedure would result in much sounder decisions than are made at the present time on the part of both the student and the college.

At a recent conference on testing for guidance, many of the test experts present favored delaying the administration of tests to determine specific aptitudes for career choices until the 12th grade. The attitude seemed to be, "If you can't provide precise predictive data which has been fully validated, delay doing anything as long as possible." The point of view of this discussion is, "If the tests and procedures are deficient, start as early as possible and supplement and check on the findings for a particular student by all available means."

Summary

What, then, is proposed as an immediate program for testing for college admission and scholarship programs?

1. Develop working statements of objectives in terms of the desirable behaviors of adults for colleges and universities. These should be focused on all aspects of life both in and after college.
2. In the absence of empirical follow-up data, use these working statements in terms of behavior for establishing policies regarding the specific aptitude, personality, and interest patterns desired in the students admitted to a college.
3. At the same time, initiate long range follow-up research programs to provide a basis for confirming or revising the tentatively established admissions and educational programs.
4. Develop tests to assist counselors in helping students formulate realistic self-concepts which have clear implications for college training and which will also assist admissions officers in deciding which applicants will contribute most to the joint program of attaining the objectives of the institution and of the individual.

The Nature of the Problem of Improving Scholarship and College Entrance Examinations

E. F. LINDQUIST, Professor of Education, State University of Iowa

Very frequently the most crucial as well as the most difficult step in solving a complex problem is that of defining the problem itself, or of clarifying the issues involved. I believe this is particularly true in relation to the problem of how to improve scholarship and college entrance examinations. It seems to me that the principal reason we have not produced more satisfactory and useful examinations of these types in the past is that we have approached the task with too narrow a concept of the problem to be solved, or with too limited a notion of the purposes such examinations should serve. Accordingly, I propose to spend most of my time here today, not in arguing the relative merits of different types of tests in relation to different purposes, but in attempting to define and clarify these purposes; and especially, in attempting to make more clear the general nature of the problem as a whole.

I propose, further, to limit my part in this discussion to college entrance and scholarship qualifying examinations that are appropriate for use in very wide-scale cooperative testing programs—programs such as the College Board or the National Merit Scholarship Qualifying Testing Programs—programs that are intended to serve a large number and wide variety of collegiate institutions and scholarship donors, as well as a highly heterogeneous and broadly inclusive population of candidates; and programs, also, in which the tests are to be administered early enough to give the candidate ample time to make his major decisions and to complete detailed arrangements for college attendance after the examination results are known. The latter means that the tests must be administered while the typical candidate is still in high school, and hence that the tests must be given by, or with the consent and approval of, the high school authorities.

It is fairly evident that nearly all college entrance and scholarship qualifying testing must be done through wide-scale cooperative programs at the high school level. It is also quite apparent that there should be only a small number of such programs. Unfortunately, this is not now the case. I do not know how many different programs of this type are now being conducted annually in the high schools of this country, but I do know that the number is far larger than can possibly be justified. It is very doubtful that the individual colleges and agencies are really any better served by these many programs than they could be by a very much smaller number properly planned on a cooperative basis. This fact is generally recognized, and as a result many high school principals are on the verge of open rebellion at what they rightly regard as the unreasonable demands made on their time and that of their pupils by this multiplicity of testing programs. This is a situation that almost but not quite calls for a natural monopoly. Some competition in the provision of test services of this kind is undoubtedly wholesome and desirable, but, for very evident and compelling practical reasons, widespread duplication of effort and consequent waste of the time of high school students and staffs must be avoided.

In order to keep this discussion within manageable limits, I propose further to consider here only examinations that are concerned with the intellectual attributes of the candidates. Other instruments or sources of information, such as interest and personality inventories, attitude scales, biographical information blanks and school records, of course, occupy a very important place in the whole process of determining college admissions or selecting scholarship recipients, but there is hardly time to consider all of these in this short discussion.

The major point that I wish to make in this paper is that our task is fundamentally one of finding a type of test that will not just serve a single well-defined purpose, but that will satisfy a fairly large number of diverse requirements. That is, the problem is one of building a *multiple* purpose rather than a single purpose test. Before attempting to identify or define these purposes and requirements individually, let us consider briefly some of the factors in the total situation that call for this multiplicity of purposes.

One of the most important of these I have already mentioned—the practical necessity of accomplishing virtually all scholarship and college entrance testing through a very small number of wide-scale cooperative programs at the high school level. The test results obtained in such programs must be useful to literally hundreds of different institutions and agencies, each of which will employ the results in somewhat different ways—often in quite markedly different ways—than most others. The results will be used by some institutions, for example, to "skim the cream" off the top of the ability distribution in a population of candidates that is already highly self-selected. In other non-selective institutions, such as state-supported universities with more applicants than they can handle, the same test results may be used to exclude students at the other end of the ability scale in the entire unselected population of high school graduates. In some institutions the results will be used to determine readiness for a broadly cultural liberal arts curriculum, or for a program of general education. In still others the results will be expected to predict success in a narrowly specialized and technical curriculum, as in colleges of engineering and the mechanical arts. In still others the same test results will be used to select students for an elementary teacher training program, or for a pre-medical course, or for a course in business management, or for a school of social welfare, and so on.

I have thus far only suggested some of the possible variations of what might be regarded as the central or most ostensible purpose of scholarship and college entrance examinations, that of selecting among the candidates on the basis of their intellectual ability. It is also extremely important to recognize that different types of tests may serve this central purpose equally well, yet may differ radically in the consequences of their use in wide-scale programs, or in their incidental and often unintended effects. For example, two types of tests may both yield scores that show the same correlation with college grades, but one may exercise a restrictive or otherwise undesirable influence on the high school curriculum and the other may not; one may foster good and the other bad attitudes towards college preparation on the part of the candidates; or one may be susceptible to superficial cramming or may lead to bad coaching practices and the other may not. In such instances, the "side effects" of the tests may often be the determining factor in test selection, and to provide for these side effects is equivalent to specifying additional purposes for the examination.

It is also extremely important to recognize that scholarship and college entrance examinations may readily be constructed so that, without any appreciable sacrifice in their ability to serve the so-called central purpose, they can serve many other equally important educational purposes as well. It is quite possible, for example, to provide a test battery that will not only predict college success or determine readiness for college as well as any other, but that will also be highly useful to high school counselors in advising students on their educational and vocational careers, or on their choice of type of college, and that will be useful as well to high school teachers in adapting instruction to individual differences, and to high school administrators in evaluating the entire educational offering of the school. Likewise, the same test battery might be useful to the college authorities for placement purposes, or for purposes of counseling and guidance, or to help them better define the college's task by more adequately describing the status and needs of their entering student body.

Planners of such testing programs should not only recognize these possibilities but should regard as one of their most important responsibilities that of thus extending the usefulness of their tests to the greatest possible extent. It is particularly important that the results obtained in scholarship testing programs be immediately useful at the high school level. The high schools are naturally reluctant to devote very much time to testing programs that are conducted solely for and in the interests of the colleges themselves. If the high schools are free to choose between two competing scholarship testing programs or services, they will undoubtedly give their support to the one in which they find the test results most useful for their own immediate purposes.

Furthermore, if we are to avoid the past mistake of conceiving too narrowly of the purposes and requirements of testing programs of this character, we must plan the programs in consideration of the present fundamental needs of the American system of education as a whole. We have been greatly concerned, recently, with the so-called Russian challenge to American education. We have become much more keenly aware of the urgent need of raising the level of intellectual competence, not only of our scientists, engineers, and technicians, but of persons engaged in all types of intellectual activities in our society. Unfortunately, the American public has been encouraged to believe that to meet this challenge, we have only to send more students to college, especially more talented students, and particularly into science and engineering courses. This has been taken to mean that increased scholarship spending, both federal and private, plus higher salaries and increased facilities for science and mathematics teaching in the public schools, is about all that is needed.

Those of our national leaders who have encouraged public belief in this apparently quick and easy and hence highly popular solution are perhaps more politicians than educators. As our educational leaders have generally recognized, our real problem is *not* how to send more students to college. We already have more students in college than our present facilities and instructional staffs will permit us to handle properly, and our present provisions for our highly talented students are especially inadequate. Our real need is not even to send more *talented* students to college. In the first place, practically all of the really highly talented students are already there. The so-called talented students who are

107

staying away from college for financial reasons are mostly fairly far down along the ability scale. If we did succeed in sending to college the very small proportion of really talented students not now in attendance, but were to do nothing more for them than we are now doing for other talented students in general, we would as a nation surely not be appreciably better off than we are at present.

Our real need, then, is not just to send a larger number of students to college, talented or otherwise, but to enrich and improve their educational experiences at all levels—college, high school, and elementary school. Our real needs are to identify the talented pupils much more surely and much earlier—long *before* they go to college—and to provide more adequately at all stages for the further development of their superior talents, and to give them the needed incentives to make the most of these enriched opportunities and of their superior abilities. According to nearly all observers of contemporary Russian education, one of the most outstanding differences between the Russian system and ours lies in the general attitude toward education, not only among students in Russian schools but among the Russian people in general, and in the interest shown in and the effort expended on self-education. If we are to meet the Russian challenge, we must, among other things, find more effective ways of motivating our students, particularly our most talented students, or of inducing them to work harder, both in and out of school, at the task of self-improvement.

While they have not generally done so in the past, wide-scale scholarship and college entrance testing programs can make a significant contribution to these basic educational needs. By providing appropriate types of examinations, the programs can give the students a concrete and immediately effective incentive to work harder at the job of getting ready for college. To serve this purpose, the examinations must measure directly the student's *readiness* for college, or the extent to which he is *prepared* to profit by the college experience. That is, they must measure as directly as possible his ability to perform exactly the same kinds of complex tasks that he will have occasion to perform in college and in his later intellectual activities in general. The examination should therefore consist in large part of exercises requiring the student to interpret and to evaluate critically the same kinds of reading materials that he will have occasion to read and study in college, and, particularly, that will require him to do the same kinds of *complex reasoning* and *problem solving* that he will have to do later both in and out of school.

If the examination is to have the maximum motivating value for the high school student, it must impress upon him the fact that his chances of being admitted to college, or of being awarded a scholarship, depend not only on his "brightness" or "intelligence" or other innate qualities or factors for which he is not personally responsible, but even more upon how hard he has worked at the task of getting ready for college, both in high school and in the years preceding high school. The examination must make him feel that he has *earned* the right to go to college by his own efforts, not that he is entitled to college admission because of his innate abilities or aptitudes, regardless of what he has done in high school. In other words, the examination must be regarded by him as an *achievement* test, or as a test of his acquired or *developed* abilities. The tasks constituting the examination must therefore obviously correspond to recognized high school learning experiences, which means that the test exercises should perhaps be grouped according to major areas of high school instruction—the social studies, the natural sciences, the humanities, the communication skills, and mathematics.

In thus organizing the tests, however, the program planners must avoid any appearance of attempting to dictate the content or the organization of the high school curriculum. The colleges must not again lay themselves open to the charge of dominating the high schools through the college entrance testing programs, as they did decades ago under the old Regents examination system in New York or under the old College Board system. The test battery may therefore *not* consist of content or subject matter examinations, corresponding to established subjects in the high school curriculum. At the same time, the importance of the student's *knowledge* must not be neglected.

These requirements can be met if the examination is concerned directly with the development of generalized intellectual skills and abilities, or with what the student can *do* with what he has learned, and if it is concerned only indirectly with *what* he has learned, in the sense of specific items of information or bits of knowledge. The necessary emphasis on the scope and quality of the student's knowledge can be secured if the exercises are based upon test situations that emphasize differences among the candidates in their general informational or ideational backgrounds, and in their previous educational experiences. That is, the exercises should give a definite advantage to the student who is already best informed *in general* about the problem to be solved, or most experienced in the solution of such problems. This can and should be done without penalizing unduly any examinee who happens not to possess a particular specific bit of information.

If the examinations are such that they provide concrete and meaningful incentives to the individual high school student, they will necessarily, and should, serve a similar purpose for the individual school as well. Examinations that measure the extent to which the individual students are prepared for college must obviously indicate also how well the schools have prepared them *for* college. The right kind of tests will therefore make the high schools more keenly aware of their own responsibilities and shortcomings, and at the same time will give them positive aid in meeting these responsibilities, by drawing their attention to broad areas or aspects of achievement most in need of improvement. I need hardly point out that college entrance examinations of the type generally regarded as intelligence tests or scholastic aptitude tests, or differential aptitude tests, are almost wholly useless for these purposes, as they are for motivating the individual student.

To appreciate fully what kind of a problem we are here considering, we must give some attention to still one more general purpose or requirement of college entrance and scholarship examinations. It seems to me that one of the most significant observations that one can make concerning such examinations is that they, more than any other single thing, constitute the real answer to the question, "Who may go to college?" To the high school student who asks, "What must I be like, or what must I be able to do to be admitted to college?" the realistic answer must be: "Different colleges have many different requirements, but there is one thing that nearly all of them require in common, there is one thing about which you may be certain—you must be able to pass the entrance examinations; that is, you must be able to do the kind of things called for by the examination questions."

If we are to look for the most universal and the really *functional definition* of the desirable college student, or of his desired intellectual attributes, we must look at the entrance examination. We should be able to find there a representative sample of precisely the same kinds of complex tasks that the college student and intellectual worker in general must be able to perform. We should expect to find there a highly meaningful definition of the things that the high school and elementary school should have prepared the student to do. I need hardly point out that, viewed as such definitions, most scholarship and college entrance examinations used in the past have been utterly inadequate.

This issue is becoming particularly important with the increased use of entrance examinations by non-selective institutions to determine who must be *denied* admission because of the institution's limited facilities. There is real danger that such institutions will place undue and uncritical reliance on entrance examinations simply because such examinations provide a demonstrably impersonal and objective basis for making unpopular decisions about the applicants. The planners of college entrance testing programs must recognize fully the responsibilities that they thus assume, and must make special efforts to provide in test form the best possible definition of what is really wanted in the incoming student.

This leads to my final general observation concerning the nature of the problem that we are here considering. It is simply that the problem is one which calls fundamentally for a *rational* rather than for an *empirical* solution. In the past, developmental work on scholarship and college entrance examinations has in general been dominated by the empirical or experimental approach. The core, if not the whole, of the examination battery has usually consisted of tests of the so-called scholastic aptitude type. In constructing such tests, we have been obsessed by the single notion that the tests must predict college success. Since we have had readily available only one quantitative measure of such success, the grade-point average, we have allowed our test selection procedures to be dominated by this dubious criterion. In many instances, tests and items have been selected almost exclusively in terms of their correlations with the grade-point averages. Furthermore, and I think this is much more significant than has been generally recognized, we have characteristically made up our experimental try-out batteries of very *short* sub-tests. In thus doing, we have placed an unduly high premium upon the reliability of the subtests, rather than upon their intrinsic validity. Our statistical techniques of test selection have tended to exclude tests of highly complex character—tests that in consequence are almost inevitably low in reliability per unit of testing time, and that for this reason alone will in short forms show low correlations with the criterion, even though their intrinsic validities are quite high. We seem to have been unduly concerned with the *efficiency* of our prediction instruments. That is, we seem to have had as our objective that of securing a high correlation with the criterion in the shortest possible amount of testing time, rather than that of attaining the highest possible validity in whatever amount of time is needed to do the job right.

What is much more serious, however, is that in selecting or constructing the sub-tests to be tried out experimentally for use in such batteries, we have strongly favored tests that are highly homogeneous in character and simple in structure, tests that show a high correlation with the criterion and low correlations with one another. That is, we have tended to exclude *complex* types of tests even from initial consideration, and often have not even tried them out in the experimental batteries. When we have included them, we have not made them long enough to compare favorably in *reliability* with the other tests tried out, and hence have doomed them to elimination later. Obviously, the final battery of tests selected cannot be any better than the tests from which the selection was made. In our efforts to analyze complex mental processes into simple and independent components, we have analyzed out or otherwise excluded the most important components of all—the most essential and distinguishing characteristic of each of which is its very complexity.

It is important, of course, that the student possess many specific skills, as well as that he have a large store of specific information. What is much more important however, is that he be able to *use* all of these skills and knowledges at the same time and in the right combination with one another in the solution of highly complex problems. It is the complex organization of, and the interactions among, these skills *as they are being used* that is most important, not just the specific skills themselves. It is extremely important, furthermore, that the student have acquired a sound sense of values, and that he be able to make decisions and to reach major conclusions in proper consideration of these values. In other words, it is most important that he exercise sound judgment in all that he does. These abilities to use many specific skills at the same time and in the right combination, to weigh values, to do complex reasoning, to exercise judgment, and many other similar abilities definitely cannot be measured in the abstract or in isolation from one another. Certainly they are not traits that are psychologically simple in structure. They can be measured, but only in relation to the complex situations in which they are demanded.

It is my contention, then, that even for the single purpose of predicting the questionable grade-point average criterion, the techniques that we have been employing are far from perfect. As a means of selecting tests that will predict multiple and complex criteria such as we should be developing to take the place of the grade-point average, they are clearly much less adequate. As a means of helping us *define* major educational goals, they are utterly inadequate.

The latter, as I see it, is the real nature of the job of improving college entrance and scholarship testing programs. I have suggested quite a number of different requirements that I believe such examinations should serve. Even so, there are many others that I have not even had time to mention. It is obviously impossible to construct a single test battery that will serve all of these many purposes perfectly, or, for that matter, that will serve any one of them to our complete satisfaction. It is possible, however, to provide a single battery that will prove *highly useful* in relation to every one of these purposes. This can be done without appreciably sacrificing the usefulness of the battery for any single purpose, including that of the prediction of success. Indeed, there is now plenty of evidence that tests of the kind that I have suggested will do a better job of predicting college success than will any available batteries of tests of specific aptitudes or skills. If we are to provide and to continue to improve the kind of tests needed, we must recognize that the task is essentially one of describing or *defining*, in terms of test situations, some of the broad educational goals that the students and the schools are now trying to attain.

I cannot emphasize too strongly, however, that it is definitely not the function of planners of scholarship and college entrance examination programs to determine or to *set* any education goals. Certainly it is not their province to attempt through the tests to *bring about changes* in the high school curriculum, no matter how desirable. It is proper and desirable, however, that the tests accurately describe some of the broad educational goals which are already universally accepted, and that they emphasize the need for further development of generalized abilities that are of self-evident and unquestioned importance. In constructing tests of this character, some statistical and empirical techniques are useful and necessary, but they are of secondary importance. This is a task that calls fundamentally for great skill, imagination, and ingenuity on the part of the item writer, rather than for skill in statistical manipulation. The writing of items for such tests calls for the very highest level of talent or competence available in the whole field of educational measurement. The whole task of test construction is one that calls for a logical rather than for an empirical approach or one that, most of all, demands the exercise of sound judgment.

In closing, I would like to repeat and emphasize a point that I made earlier—that it has not been my purpose here to argue the merits of any particular test or testing programs. My concern here is only with the general direction in which our further efforts at test *improvement* should be pointed. We have made great progress in the past in developing tests of the type suggested, but certainly there is plenty of room for and great need for continued improvement in all present instruments. I am convinced that, if we approach this task of test improvement in the manner suggested, if we recognize that scholarship and college entrance examinations can and should serve a wide variety of purposes, and if we recognize that our task is fundamentally one of defining the generalized and complex abilities we want the high school student graduate to have developed, we can build into such programs some really significant positive values for American education.

What Kinds of Tests for College Admission and Scholarship Programs?

ALEXANDER G. WESMAN, Associate Director, Test Division, The Psychological Corporation

The tests we need for college admissions purposes are those which are reliable, efficient, inexpensive, confidential, comprehensive, unique, reflective of the curriculum, independent of the curriculum, fair to late developers, and valid for every curriculum in every institution of higher learning. Unfortunately, no such set of tests exists. In fact, no such set of tests can exist. The demands of each institution are, or should be, unique—the tests one college needs will necessarily differ in some ways from the tests other institutions need. Any attempt to specify the same testing program for all colleges, or all scholarship purposes, is inherently self-defeating.

The central issue in choosing tests for college admissions purposes is the same as for any other purposes—what use is to be made of the test results? This in turn depends on the nature of the individual institution—its goals, its role in our society, its facilities, its philosophy. The growing prevalence of national and statewide programs embodies a real danger that individual differences among institutions of higher education will be overlooked. The advantages of uniform testing programs may be purchased at the excessive price of ignoring one of the greatest strengths of our educational system—the variety of functions performed by our colleges and universities. If every institution were concerned with selecting for admission only the intellectually elite and in providing the same kind of education to all those it admitted, a single set of tests might be prescribed for all. As long as we have state universities and highly selective private colleges, liberal arts colleges and agricultural colleges, cultural emphases and vocational emphases, it is unlikely that one set of tests, or one program of tests however thoughtfully devised, will adequately serve the needs of all. Rather a variety of tests and a variety of programs is essential if each institution is to approximate the requirements of its own special circumstances.

Certainly every institution needs tests which are reliable; but is the same test reliable for every institution? Not unless it is a most inefficient

114

instrument. Our institutions of higher education vary widely with respect to the levels of talent in the students they admit. If a test is to have enough difficult material to discriminate reliably among the top twenty per cent of our students, and enough easy material to discriminate reliably among the bottom twenty per cent of college freshmen, the test will be far too long for practical use in either group. Further, the large number of easy items is likely to bore the better students almost as much as the large number of difficult items frustrates the less able; and, in each case, the sections of inappropriate difficulty represent inefficient measurement.

An example may focus the problem of range of talent more sharply. In one state university system last year, the same set of tests was used in all the units of the system. The average verbal test score of the entering freshmen in one of the institutions (College A) was more than two standard deviations below the average score of those entering another institution (College B) in the same system. College B, whose freshmen scored highest in the state system, was only average among all the schools which use this test. This suggests that differences in mean scores between the most selective schools using this test and low-scoring College A are fully four to five standard deviations. It is probable that for a majority of the students in College A the later half of the test provided a depressing experience, but no real measurement. For these examinees the effective portion of the test was composed of perhaps half the items printed in the booklet. The reliability of the test for discriminating among these students must be assumed to have suffered accordingly.

The tests should be efficient—they should occupy as little of the school's and the student's time as is necessary. This is not to say that the time spent in testing is not as well spent as if equal time were devoted to other kinds of experience to which a student might be exposed. Rather, it suggests that efficient tests permit the gathering of more information within reasonable time limits. If four hours are to be devoted to testing, we should seek full value for those four hours. There are programs which compel some students to stay overnight in an out-of-town lodging; if more efficient testing can eliminate this burden, such programs should be made more efficient. The student may not be in a position to protest; but the captive state of the student should not make his captors less merciful.

To be most effective, tests should supplement information which is otherwise available, rather than duplicate such information. The college which draws its students from a small number of local secondary schools should be able to accept the students' high school records as evidence of their academic preparation; achievement tests are of secondary utility if they are devoted to assessing the same information as is represented by well-understood school grades. Where there is considerable diversity among feeder schools in their curricula and grading standards, the use of achievement tests may be more defensible. However, in our enthusiasm for tests we should not forget what research has so often demonstrated—that even where students are drawn from diverse secondary schools, high school average is often one of the best predictors of performance in college. Accordingly, instruments which are less directly reflections of the subject matter competence of the student may provide more new information concerning him than do tests in subject matter for which grades are already available.

There are at least two other advantages to the use of non-curricular tests. The use of achievement tests for college admissions all too often exerts a disproportionate influence on the secondary school curriculum and the secondary school teacher. Achievement tests are valid if they measure what the school wants to teach; but schools frequently behave as though their teaching is valid if their students do well on some esteemed achievement test. Some years ago it was commonplace in New York to hear the complaint that the final semester of a course was devoted entirely to specific preparation of the students for the Regents examination; then, for several years, the issue appeared to have been resolved. Today, once again, another set of tests occupies a similarly dominant position. Those of us who are responsible for developing such tests may find ready refuge in the statement that *we* do not recommend that the school or teacher adopt this subservient role—that the tests are intended to follow, not determine, the curriculum. But as long as subject matter tests serve college admissions purposes, we must expect teachers who are anxious to help college-bound students—and teachers whose own performance will be judged by their pupils' success on these tests—to concentrate on the tests as much as on the course.

A second advantage of tests which are not curriculum oriented is their potential for rescue functions. There are students whose formal academic preparation is defective—those who were unstimulated by their courses or their teachers—who may nevertheless be salvaged. Their previous failure to learn may have been the result of delayed maturity on their part, or of an uninspired educational environment. That these students have *not* learned what their courses offered would be documented by course grades and by achievement tests alike. To reveal that they *could* learn requires a different kind of predictive measure.

The proportion of students who do poorly in high school, then find themselves when they are given a chance to do college work, may be small, but the absolute number of such students is large enough to warrant serious attention. A large midwestern state university found in its 1957 entering class 188 students who scored among the top twenty-five per cent on the *College Qualification Tests*, and were in the lowest quarter of their high school class. More than half of these students attained a first semester grade point average of 2.0 (C) or better. In this same class, there were 314 freshmen who were also in the top quarter on the CQT and in the third quarter in high school rank. Three-fourths of these freshmen earned a grade point average of 2.0 or better. These are students—300 of them in a single freshmen class—for whom prognosis on the basis of past academic achievement would be pessimistic, but who were correctly identified by the tests as being capable of at least initially satisfactory work in college.

There are a number of considerations which each institution must resolve for itself before it adopts an admissions battery or accepts a program devised by some outside agency.

1. Will a policy of selective admissions be practiced, or is the institution obliged, perhaps by state charter, to admit all applicants who meet certain minimum requirements?

A highly esteemed private liberal arts college has the privilege of selecting only those students who show greatest intellectual promise. Some state universities do not have that same privilege. Society has prescribed that not only the elite shall be educated; all who can profit from collegiate education in diverse curricula and at varying levels of intellectual demand are to be given the opportunity for further academic training. It is true that even publicly supported institutions are finding it necessary, because of swelling hordes of applicants and limited classroom capacities, to exercise some selection. But the exclusion of the least promising from the great mass of applicants is a quite different task from that of choosing a small number of the elite from an already self-selected group of top-ranking candidates. One should expect that tests differing in difficulty, and perhaps in kind, are necessary for these different tasks.

2. Whether or not selective admissions will be practiced, are students placed in different classes or sections on the basis of test results?

If freshman courses are offered at more than one level to students with different academic preparation, achievement tests may be useful to appraise the student's competence at entrance. If there is a course in chemistry for advanced students and another for students who have not previously taken chemistry—and *if* the student's record of high school courses is judged insufficient to testify to his knowledge of the subject—a chemistry test may be advisable. If, on the other hand, the same first course is offered to all students regardless of previous exposure, a subject matter test is probably less crucial.

3. Will the faculty make use of the test results in its teaching, or are the tests primarily to serve a screening function?

If the biology or history teacher will use the information gained from tests in his subject matter to plan his instruction, achievement tests may be desirable for students entering the course. There is many a faculty member who, rightly or wrongly, expresses indifference to how much subject matter the student has learned before he enters the class; rather, it is whether the student can learn, and is willing to learn, what the professor wishes to teach him that is crucial. This teacher may be one who is unconvinced by the suggestion that what the student has learned in the past is predictive of what the student will learn in the future. Or, this may be a teacher who, faced by overflowing classes of students with a wide range of previous preparation in the subject, has recognized the futility of trying to tailor his teaching to the varying amounts of knowledge possessed by the individual entering freshmen. All students are treated by these two professors as essentially equal and uniformed on entrance to the class. High school course records in the subject are not seen as helpful by these professors; achievement test results are likely to be equally ignored by them.

If a test is to be used primarily as a screening device rather than as a basis for instruction, a scholastic aptitude measure will be more efficient and more broadly applicable than a subject matter test.

4. How many curricula does the school offer?

The small liberal arts college may require that all students take a standard, prescribed curriculum for the most part, with a small number of electives. A large state university may be composed of a number of colleges—such as science and letters, agriculture, education, nursing, pharmacy and engineering—with each college in turn offering more than one curriculum. The variety of curricula in the state university probably assures that the student bodies of the several colleges also vary, in level of ability as well as in areas of academic interest. A minimal program which would prove satisfactory for the homogeneous freshman class of the liberal arts college might well prove inadequate for the heterogeneous population which enters the complex state university.

Dr. Harold Gulliksen, discussing several papers at a recent symposium said, in effect, "Once again we have heard excellent presentations of the problems in this area and the questions that need to be asked. It would be nice if, sometime, we might have a paper which supplied answers." Since the views embodied above express the conviction that no single battery of tests will serve equally in all institutions, an attempt to propose such a battery would be inconsistent. Nonetheless, a general approach can be presented.

The core of an admissions testing program should include measures which appraise the student's command of our two most important symbol systems, verbal and quantitative. The ability to manipulate verbal and numerical concepts has almost invariably been shown to be associated with success in future learning at all educational levels. Opinions differ as to how these abilities may best be tapped—by synonyms, antonyms or verbal analogies, by number series, problem-solving, or numerical computation—but there should be little dispute that some kind of effective appraisal of verbal and numerical abilities is essential. A third component might be a brief test of information, sampling broadly from the general areas of physical and social science. This test would be intended to provide some reflection of the student's educational background where feeder schools and their marking systems are diverse. As a fourth component, a reading test might well be included as much for use in guidance and identification of students in need of remedial training as for predictive purposes. Then, because the student's outlook toward school may indicate how he will react to the educational process, a survey of his beliefs and attitudes with respect to study, to teachers and to the general academic environment might well be in order. Beyond this core, additional testing with respect to special abilities (such as space perception) or specific subject matter competence (e.g., formal mathematics), may be added according to the particular character of the institution and the readiness of faculty to utilize the test results as a basis for teaching.

With respect to tests as a basis for awarding scholarships, one needs first to inquire what purpose the scholarships are to serve. If we are simply seeking the academically most promising, then a verbal and numerical test of sufficient difficulty to challenge the top two, or five, or ten per cent of secondary school graduates will do a satisfactory job. If scholarships are intended to provide additional recruits for special areas of our society—scientists, social workers, teachers or missionaries—the tests to be used, and scores which will qualify the accepted candidates, must be tailored to the task.

One could make a brief for examining the subject area competence of students in areas in which they had not had previous schooling. There are undoubtedly students who have learned a great deal about mechanical things through their own curiosity, through recreation and experimentation, through observation and self-directed reading rather than through formal course work. If scholarships encouraged the further development of students such as these, a potential additional source of creative talent might be uncovered. A brief could be made, too, for awarding scholarships to those who are not in the highest ranks of academic promise but who can contribute importantly nonetheless. We award scholarships to students many of whom would go on to college in any event. If instead, potential teachers could be located and subsidized—students who are not in the top ten per cent of their class, whose scores on our usual scholarship tests are mediocre, who would not otherwise pursue further education, who might not earn the highest grades in a teacher training course but who could successfully negotiate a teachers college program and would enjoy teaching—more would be contributed by scholarships for this purpose than by the ego-satisfying but unessential support of those whose careers are not genuinely affected. Lower qualifying scores, or even different examinations, should be employed for this kind of scholarship award.

To summarize: the kinds of tests that are appropriate for college admission and scholarship programs are those which are best suited to the individual institution and the particular purposes of the scholarship donor. No one testing program will suit all schools or all purposes. There are many good tests. It is incumbent on the conscientious user to select from among them those which most nearly meet his special needs and circumstances. Otherwise, the tests which have provided milestones along the road of educational progress may become millstones around the neck of the educational process.

Appendix

Participants—Invitational Conference on Testing Problems

Adkins, Dorothy C., University of North Carolina
Adragne, Clara, Palm Beach, Florida
Albitz, Diane R., Alexandria, Virginia
Alman, John E., Boston University
Almy, Millie, Teachers College, Columbia University
Alt, Pauline M., Teachers College of Connecticut, New Britain, Connecticut
Anastasi, Anne, Fordham University
Anderson, Gordon V., University of Texas
Anderson, Howard R., University of Rochester
Anderson, Rose G., New York City
Anderson, Scarvia B., Educational Testing Service
Anderson, T. W., Columbia University
Andree, Robert G., Rich Township High School, Park Forest, Illinois
Andrews, R. Duane, Wyoming State Department of Education
Andrews, T. G., University of Maryland
Angoff, William H., Educational Testing Service
Armstrong, Fred, U. S. Steel Foundation
Aronow, Miriam S., New York City Bureau of Educational Research
Atkins, William H., Rutgers University
Ayars, Albert L., Hill and Knowlton Inc.
Baier, Donald E., General Electric Co.
Bannon, Charles J., Waterbury (Connecticut) Department of Education
Bardack, Herbert D., New York State Department of Civil Service
Barker, Roland, Montclair (New Jersey) Public Schools
Barnes, Paul J., World Book Company
Barry, Ruth E., Hunter College, New York City
Bartnik, Robert V., Educational Testing Service
Bayroff, A. G., Department of the Army
Becker, Theodore, New York State Department of Civil Service
Bedard, Joseph A., New Britain (Connecticut) Public Schools
Bell, Lowell, South Dakota Department of Public Instruction
Belt, Sidney L., Educational Testing Service
Bement, Dorothy M., Northampton (Massachusetts) School for Girls
Bennett, George K., The Psychological Corporation
Bennett, Mrs. George K., The Psychological Corporation
Bennett, Ralph, New York City
Benson, Arthur L., Educational Testing Service
Bentzen, Frances, Syracuse University
Berger, Bernard, New York City Department of Personnel
Bergesen, B. E., Jr., Personnel Press, Inc., Princeton, New Jersey
Birnbaum, Allan, Columbia University
Bixler, H. H., Western Carolina College, Cullowhee, North Carolina
Bligh, Harold F., World Book Company
Bloom, Benjamin S., Univ. of Chicago
Bloomer, Richard H., State University Teachers College, Geneseo, New York
Bollenbacher, Joan, Cincinnati (Ohio) Public Schools
Braca, Susan E., Oceanside Senior High School, Forest Hills, New York
Brandt, Hyman, American Occupational Therapy Association, New York City
Bristow, William H., Bureau of Curriculum Research, New York City
Brochard, John H., University of Buffalo

Broderick, J. Lawrence, Y.M.C.A., New York City
Brolyer, Cecil R., New York State Department of Civil Service
Brooks, W. Douglas, Educational Research Corp., Cambridge, Mass.
Brown, David, Educational Testing Service
Brown, Fred S., Great Neck (New York) Public Schools
Brown, Leanna, Educational Testing Service
Brown, F. Martin, Fountain Valley School, Colorado Springs, Colorado
Bryan, J. Ned, North Central Association, Superior and Talented Student Project, Chicago, Illinois
Bryan, Joseph G., American Machine & Foundry Co.
Bryan, Miriam M., Rutgers University
Buckingham, Guy E., Allegheny College, Meadville, Pennsylvania
Burdock, E. I., Biometrics Research, New York City
Burke, James M., Darien (Connecticut) Public Schools
Burke, Paul J., Bell Telephone Labs
Burnham, Paul S., Yale University
Buros, Luella, Rutgers University
Buros, Oscar, Rutgers University
Calhoun, Mrs. John D., Fieldston School, New York City
Capehart, Bertis E., Hill and Knowlton, Inc.
Capps, Marian P., Virginia State College, Norfolk, Virginia
Carroll, John B., Harvard University
Carstater, Eugene D., Bureau of Naval Personnel
Cashman, Jerome P., Archdiocesan Vocational Service, New York City
Cayne, Bernard S., Ginn and Company
Chacko, Yo Yee, New York City Board of Higher Education
Chappell, Bartlett E. S., New York Military Academy
Chauncey, Henry, Educational Testing Service
Christensen, Clifford M., New York State Department of Education
Clark, Pamela, Educational Testing Service
Clark, Ronald J., St. Paul's School, Concord, New Hampshire
Cleary, Robert, Educational Testing Service
Clendenen, Dorothy M., The Psychological Corporation
Cliff, Norman, Educational Testing Service
Clifford, Paul I., Atlanta University
Cobes, Jeanne M., Educational Testing Service
Coffield, William, Alabama Polytechnic Institute, Auburn, Alabama
Coffman, William E., Educational Testing Service
Cole, Joseph W., University of Rochester
Coleman, Elizabeth R., Vocational Education and Extension Board, Rockland County, New York
Conrad, Herbert S., U. S. Office of Education
Cook, Desmond L., Purdue University
Cory, Charles H., Philadelphia Department of Personnel
Cosgrove, John E., New York State Department of Civil Service
Crane, Harold L., Jr., Educational Testing Service
Craven, Ethel Case, Polytechnic Institute of Brooklyn
Crawford, J. R., University of Maine
Crawford, Kay E., New York City Board of Higher Education
Crook, Frances, McGill University
Crootof, Charles, Martin Van Buren High School, New York City
Crosson, Wilhelmina M., Palmer Memorial Institute, Sedalia, North Carolina
Cummings, Mary B., Boston (Massachusetts) Public Schools
Dailey, John T., American Institute for Research, Washington, D. C.
Damrin, Dora E., Educational Testing Service
Davidson, Helen H., The City College of New York

PARTICIPANTS

DAVIS, Anne D., Educational Testing Service
DAVIS, Frederick B., Hunter College, New York City
de JUNG, John, Department of the Army
DENISON, Violet, Educational Testing Service
DERWIN, Edward P., Crosby High School, Waterbury, Connecticut
DIAMOND, Esther E., Science Research Associates
DIAMOND, Lorraine K., Teachers College, Columbia University
DIAMOND, M. David, Riverside Hospital, New York City
DICKSON, Gwen, National Cathedral School, Washington, D. C.
DIEDERICH, Paul B., Educational Testing Service
DIGGS, Franklin B., New York City Department of Personnel
DION, Robert, California Test Bureau
DOBBIN, John, Educational Testing Service
DOPPELT, Jerome E., The Psychological Corporation
DOWNES, Margaret, New York State Department of Civil Service
DRAGOSITZ, Anna, Educational Testing Service
DRY, Raymond J., Life Insurance Agency Management Association
DUBIN, S. S., Pennsylvania State Department of Health
DUBNICK, Lester, Municipal Colleges of the City of New York
DUDEK, E. E., U. S. Naval Personnel, Research Field Activity
DUFF, Franklin L., World Book Company
DUKER, Sam, Brooklyn College
DULAC, Frances, Niskayuna Senior High School, Schenectady, New York
DUNN, Frances E., Brown University
DUROST, Walter N., Board of Public Instruction, Largo, Florida
DYER, Helen S., Princeton, New Jersey
DYER, Henry S., Educational Testing Service
EBEL, Robert L., Educational Testing Service
EDDY, Robert P., Rutgers University
EDELSTEIN, J. David, New York City Study of Mentally Handicapped Children
EDGERTON, Harold A., Richardson, Bellows, Henry & Co., Inc.
EDMISTON, Andrew J., Lehigh University
EDWARDS, Winifred, Irvington (New Jersey) High School
EELLS, Kenneth, University of Illinois
ENGELHART, Max D., Chicago City Junior College
EPSTEIN, Bertram, The City College of New York
FALCO, Rose M., World Book Company
FATERSON, Hanna, New York City
FAY, Paul J., New York State Department of Civil Service
FELDT, Leonard S., State University of Iowa
FENDRICK, Paul, Western Electric Company
FENOLLOSA, George M., Houghton Mifflin Company
FENSTERMACHER, Guy M., Educational Testing Service
FERGUSON, John P., The Pingry School, Elizabeth, New Jersey
FERGUSON, Leonard W., Life Insurance Agency Management Association
FERRIS, Anne H., Educational Testing Service
FERRIS, Frederick L., Jr., Educational Testing Service
FIELDS, Harold, New York City Board of Education
FIFER, Gordon, Hunter College, New York City
FINDLEY, Warren G., Atlanta (Georgia) Board of Education
FINEGAN, Owen T., Gannon College, Erie, Pennsylvania
FINK, David R., Jr., University of Maine
FISHMAN, Joshua A., Greenfield Center for Human Relations, Philadelphia
FLANAGAN, John C., American Institute for Research, Pittsburgh
FLEISCH, Sylvia, Boston University
FLEMMING, Edwin G., Burton Bigelow Organization, New York City

Fontaine, Thomas, National Science Foundation
Forrester, Gertrude, West Side High School, Newark, New Jersey
Fort, G. H., Atlanta (Georgia) Board of Education
Foshay, Arthur W., Teachers College, Columbia University
Freas, Howard J., Jr., Hatboro-Horsham High School, Philadelphia
Freeman, Paul M., Educational Testing Service
French, Benjamin J., New York State Department of Civil Service
French, John W., Educational Testing Service
Fricke, Benno G., Univ. of Michigan
Friedman, Sidney, Bureau of Naval Personnel
Fruchter, Ben, University of Texas
Frutchey, Fred P., U. S. Department of Agriculture
Gay, George, Jewish Vocational Service, Milwaukee, Wisconsin
Gallagher, Henrietta L., Educational Testing Service
Georgia, Sister M., Rosary Hill College, Buffalo, New York
Gerberich, J. Raymond, University of Connecticut
Gilbert, Arthur C. F., Princeton University
Gilbert, Harry B., New York City Board of Education
Glaser, Robert, University of Pittsburgh
Glickman, Albert S., U. S. Naval Personnel, Research Field Activity
Gobetz, Wallace, New York University
Godshalk, Fred I., Educational Testing Service
Goldberger, Frances R., New York City Department of Personnel
Goldman, Leo, Brooklyn College
Goldstein, Leo S., Cornell University Medical College
Goodman, Samuel M., New York State Department of Education
Gooler, Edward, Central Islip State Hospital, East Islip, New York
Gordon, Leonard V., U. S. Naval Personnel, Research Field Activity
Gotkin, Lassar, Teachers College, Columbia University
Gray, Mrs. Lyle Blaine, Baltimore, Maryland
Greene, Edward B., Chrysler Corporation
Grinsted, Alan D., Department of the Navy
Grudel, Regina C., Southbury (Connecticut) Training School
Guerriero, Michael A., The City College of New York
Guilford, J. P., University of Southern California
Gulliksen, Harold, Educational Testing Service
Gustad, John W., University of Maryland
Guthrie, George M., Pennsylvania State University
Haagen, C. Hess, Wesleyan University
Hadley, Everett E., West Hartford (Connecticut) Public Schools
Hagin, Rosa A., Irvington (New Jersey) Public Schools
Hagman, Elmer R., Greenwich (Connecticut) Public Schools
Hall, Robert G., Manter Hall School, Cambridge, Massachusetts
Hardesty, Mrs. F. P., Biometrics Research, New York City
Harmon, Lindsey R., National Academy of Sciences
Harvey, Philip R., Educational Testing Service
Hastings, J. Thomas, Univ. of Illinois
Hausman, Howard J., National Science Foundation
Hayes, M. Joyce, University of North Carolina
Hayward, John C., Bucknell University
Healy, Ernest A., Jr., Center for Psychological Service, Washington, D. C.
Heaton, Kenneth L., Heaton, Floyd and Watson, Philadelphia
Heinkele, Francis W., Pace College, New York City
Helm, Carl E., Educational Testing Service

PARTICIPANTS

HELME, William H., Department of the Army

HENRY, Sallyann, The Psychological Corporation

HERRICK, C. James, Rhode Island College of Education

HICKS, Ernest H., Jersey City State College

HIERONYMUS, A. N., State University of Iowa

HILLS, John R., Regents of the University System of Georgia

HITTINGER, William F., University Research Associates, State College, Pa.

HOLLIS, Esther, The Psychological Corporation

HOPMANN, Robert P., Board for Higher Education, Missouri Synod of the Lutheran Church

HOROWITZ, Leola S., Adelphi College, Garden City, New York

HOROWITZ, Milton W., Queens College, Flushing, New York

HORTON, Clark W., Dartmouth College, Hanover, New Hampshire

HUDDLESTON, Edith M., Washington, D. C.

HUDEN, John C., University of Vermont

HUNT, Thelma, George Washington University

HYPPS, Irene C., Public Schools of the District of Columbia

JACKSON, Douglas N., Pennsylvania State University

JACOBS, Robert, International Cooperation Administration, Washington, D.C.

JANEBA, Hugo B., Rutherford (New Jersey) High School

JASPEN, Nathan, National League for Nursing, Inc., New York City

JOHNSON, A. Pemberton, Newark (New Jersey) College of Engineering

JOHNSON, Genevieve G., Niskayuna Senior High School, Schenectady, New York

JONES, Lyle V., University of North Carolina

JOSLOW, Samuel, The Reading Institute of Boston

KABACK, Goldie Ruth, The City College of New York

KALMBACH, R. Lynn, Columbia (South Carolina) City Schools

KAPLAN, Bernard A., New York State Department of Education

KATZ, Martin R., Educational Testing Service

KEEFE, John M., New York State Department of Civil Service

KELLER, Franklin J., Community Talent Search, National Scholarship Service & Fund for Negro Students, N. Y. C.

KELLEY, H. Paul, University of Texas

KELLEY, Mrs. H. Paul, Austin, Texas

KHOO, Eng Choon, Educational Testing Service

KING, Richard G., Harvard College

KIRKPATRICK, Forrest H., Bethany College, Bethany, West Virginia

KLINE, William E., Baltimore (Maryland) Board of Education

KOCH, John C., Jr., Millburn (New Jersey) Junior High School

KOSTICK, Max Martin, State Teachers College at Boston

KURTZ, Albert K., The Psychological Corporation

KUSHNER, Rose E., Teachers College, Columbia University

KVARACEUS, W. C., National Education Association Juvenile Delinquency Project

LADD, Carl, Science Research Associates

LADUKE, Charles V., U. S. Armed Forces Institute, Madison, Wisconsin

LAMKE, Thomas A., Iowa State Teachers College

LANE, Frank T., State University of New York

LANG, G., Fairleigh Dickinson University, Teaneck, New Jersey

LANGMUIR, C. R., The Psychological Corporation

LANKTON, Robert S., Detroit Board of Education

LANNHOLM, Gerald V., Educational Testing Service

LEBOLD, W. K., Purdue University

Lebow, Daniel B., New York City Department of Personnel
Lee, Judith, New York City
Lennon, Roger T., World Book Company
Levine, Richard, Educational Testing Service
Lindquist, E. F., State University of Iowa
Litterick, W. S., The Harley School, Rochester, New York
Loevinger, Jane, Jewish Hospital, St. Louis, Missouri
Logue, John, Philadelphia Department of Personnel
Long, Lillian D., Professional Examination Service, New York City
Long, Louis, The City College of New York
Long, W. F., U. S. Air Force Academy, Denver, Colorado
Lord, Frederic, Educational Testing Service
Lorge, Irving D., Teachers College, Columbia University
Loughran, George A., St. Francis College, Brooklyn, New York
Lutz, O. M., Montclair (New Jersey) State College
Lyons, William A., New York State Department of Education
MacGinitie, Walter H., Teachers College, Columbia University
Mangold, Eugene F., Jesuit Educational Association, New York City
Manoil, Adolph, Park College, Parkville, Missouri
Manwiller, Charles E., Pittsburgh Public Schools
Marie, Sister Margaret, Trinity College, Washington, D. C.
Marston, Helen M., Educational Testing Service
Mathews, Anne, Educational Testing Service
Mathews, Chester O., Ohio Wesleyan University
Mathewson, Robert H., New York City Board of Higher Education
Mattioli, Michael J., New York State Department of Education
Mayhew, Lewis B., Michigan State University
McCabe, Frank J., Raytheon Manufacturing Company
McCall, W. C., University of South Carolina
McCann, Forbes E., Philadelphia Department of Personnel
McCarthy, Dorothea, Fordham University
McCracken, Charles W., Trenton (New Jersey) State College
McFee, Anne, Syracuse University
McGuire, Anne, New York State Department of Social Welfare
McIntire, Paul H., University of New Hampshire
McLain, Joseph C., Mamaroneck (New York) High School
McLaughlin, K. F., U. S. Office of Education
McManus, Anna L., Harvard University
McManus, Leo F., Jr., University of Connecticut
McNamara, W. J., International Business Machines Corporation
McNemar, Quinn, Stanford University
McTernan, John W., State University Teachers College, Plattsburgh, New York
Mead, Harriet, Metropolitan Life Insurance Company
Meade, Martin J., Fordham University
Medley, Donald M., Municipal Colleges of New York City
Meltzer, Elihu M., New York City Department of Personnel
Melville, S. Donald, Educational Testing Service
Mernyk, Charlotte L., Brooklyn, New York
Merwin, Jack C., Syracuse University
Messick, Samuel, Educational Testing Service
Metz, Elliott, New York University
Michael, William B., University of Southern California
Milholland, John E., University of Michigan

PARTICIPANTS

MILLER, Harold T., Houghton Mifflin Company
MITCHELL, Blythe C., World Book Company
MITTMAN, Arthur, State University of Iowa
MITZEL, Harold E., Municipal Colleges of New York City
MOLL, Clarence R., Pennsylvania Military College, Chester, Pennsylvania
MOLLENKOPF, William G., Procter & Gamble
MOORE, James W., California State Scholarship Commission
MORRIS, John B., University of Mississippi
MORRISON, Alexander W., Polytechnic Institute of Brooklyn
MOSELY, Russell, Wisconsin State Department of Education
MUIRHEAD, Peter, U. S. Office of Education
MURRAY, John E., U. S. Naval Training Device Center, Port Washington, New York
MYERS, Charles T., Educational Testing Service
MYERS, Sheldon S., Educational Testing Service
NALLY, Thomas, University of Rhode Island
NEVIN, Margaret, Educational Testing Service
NOLL, Victor H., Michigan State University
NORTH, Robert D., Educational Records Bureau
NOSOW, Sigmund, Michigan State University
NUGENT, William F., Jersey City State College
OATHOUT, D. V., American Telephone & Telegraph Company
O'BRIEN, Brian, Fordham University
ODGERS, John G., Ohio Scholarship Tests, Columbus, Ohio
OLSEN, Marjorie, Educational Testing Service
O'NEILL, Helen, World Book Company
ORLEANS, Joseph B., George Washington High School, New York City
ORSHANSKY, Bernice, New Rochelle (New York) High School
ORR, David B., American Institute for Research, Washington, D. C.
PACE, C. Robert, Syracuse University
PACKARD, Albert G., Baltimore (Maryland) Public Schools
PAGCALIWAGAN, Luciana C., State University of Iowa
PALMER, Orville, Educational Testing Service
PARRES, John G., Delaware State Department of Public Instruction
PARTINGTON, Dorothy, Educational Testing Service
PASHALIAN, Siroon, N. W. Moody Corporation
PEARSON, Richard, College Entrance Examination Board
PERLOFF, Robert, Science Research Associates
PERRY, W. D., University of North Carolina
PETERS, Frank R., Ohio State University
PETERSON, Donald A., Life Insurance Agency Management Association
PHILPOTT, Emily L., Queens College, Flushing, New York
PIERSON, George A., Queens College, Flushing, New York
PITCHER, Barbara, Educational Testing Service
POLLACK, Norman C., New York State Department of Civil Service
PRATT, Carroll C., Princeton University
PRESTON, Braxton, Educational Testing Service
QUINN, John S., Jr., World Book Company
RABINEAU, Louis, Pratt Institute, Brooklyn, New York
RABINOWITZ, William, Municipal Colleges of the City of New York
RANDALL, Harvey, New York State Department of Civil Service
RAPHAEL, Brother Aloysius, Bishop Loughlin Memorial High School, Brooklyn, New York
RAPPARLIE, John H., Owens Illinois Glass Company

Raskin, Evelyn, Brooklyn College
Regan, James J., U. S. Naval Training Device Center, Port Washington, New York
Reicker, Ronald, The Pingry School, Elizabeth, New Jersey
Reilly, J. J., St. Anselm's College, Manchester, New Hampshire
Reiner, Harry, New York City Department of Personnel
Remmers, H. H., Purdue University
Ressler, Helene, Educational Testing Service
Rethlingshafer, Dorothy, University of Florida
Rhodes, David, Society for the Investigation of Human Ecology, Washington, D. C.
Ricciuti, Henry N., Cornell University
Richter, Virginia B., Childrens' House, Garrett Park, Maryland
Ricks, J. H., Jr., The Psychological Corporation
Righthand, Herbert, J. M. Wright Technical School, Stamford, Connecticut
Rim, Y., New York University
Rimalover, Jack K., Educational Testing Service
Buddeke, Sister Rita, Catholic University of America
Ritenour, Louise R., Educational Testing Service
Rivlin, Harry N., New York City Board of Higher Education
Roblyer, William A., The Choate School, Wallingford, Connecticut
Roca, Pablo, Puerto Rico Department of Education
Roehrig, Albert K., Phillips Academy, Andover, Massachusetts
Romanella, Alan, Board of Cooperative Educational Services, West Sayville, N. Y.
Rosenblatt, Nancy, Educational Testing Service
Rosenthal, Elsa, Educational Testing Service
Rosinski, Edwin F., University of Buffalo
Rosner, Benjamin, Rutgers University
Roth, Robert, Hampton Institute, Hampton, Virginia
Rulon, Paul J., Harvard University
Rummel, J. F., University of Oregon
Runkel, Philip J., University of Illinois
Ryans, David G., University of Texas
Sait, Edward, Rensselaer Polytechnic Institute
Salter, Katharine A., New York University
Samler, Joseph, Veterans Administration, Washington, D. C.
Sanders, Edward, Pomona College, Claremont, California
Satterfield, M. L., New York State Department of Civil Service
Saunders, David R., Educational Testing Service
Savitt, Arthur H., New York City Police Academy
Scates, Alice Y., U. S. Office of Education
Schapiro, Harold B., Young & Rubicam, Inc., New York City
Schein, Abraham, New York City Department of Personnel
Schepers, J., Educational Testing Service
Schneider, Richard, New York State Department of Civil Service
Schrader, William B., Educational Testing Service
Schroedel, E. C., International Business Machines Corporation
Schulman, Frances H., Institute for Career Guidance, New York City
Schulman, Jay, Institute for Career Guidance, New York City
Schweiker, R. F., Educational Research Corporation, Cambridge, Mass.
Scott, C. Winfield, Rutgers University
Scott, Winifred S., New Brunswick, New Jersey
Seashore, Harold, The Psychological Corporation
Segel, David J., Indiana University
Seibel, Dean W., Educational Testing Service
Setzer, Charles J., New York City Department of Personnel

PARTICIPANTS

Sforza, Richard F., New York State Department of Civil Service
Shanner, W. M., California Test Bureau
Sharp, Catherine G., Educational Testing Service
Shaycoft, Marion F., American Institute for Research, Washington, D. C.
Shea, James E., New York State Department of Civil Service
Shear, Bruce E., New York State Department of Education
Shields, Mary, National League for Nursing, Inc., New York City
Shields, O. L., Jefferson County (Kentucky) Schools
Shimberg, Benjamin, Educational Testing Service
Siciliano, Charles, New York State Department of Education
Silverberg, Edward, New York City Department of Personnel
Simon, Dorothy M., New York City Department of Personnel
Singleton, Carol, Bank Street College of Education, New York City
Sitgreaves, Rosedith, Teachers College, Columbia University
Smith, Alexander, F., New Haven (Connecticut) State Teachers College
Smith, Allan B., University of Connecticut
Smith, Ann Z., Educational Testing Service
Smith, Lyman J., Illinois State Scholarship Program
Smith, Marshall P., Trenton (New Jersey) State College
Smith, Robert E., Educational Testing Service
Snodgrass, Robert, Educational Testing Service
Solomon, Robert J., Educational Testing Service
Souther, Mary Tayloe, Tower Hill School, Wilmington, Delaware
Spaney, Emma, Queens College, Flushing, New York
Spaulding, Geraldine, Educational Records Bureau
Speer, George S., Illinois Institute of Technology
Spencer, W. Douglas, New York City
Spencer, Wesley G., Cambridge, Mass.
Sprague, Arthur R., Hunter College, New York City
Sprague, Margery, Ft. Lee, New Jersey
Stafford, Richard, Princeton University
Stake, Robert, University of Nebraska
Stecklein, John E., University of Minnesota
Stewart, Mary, New York University, Bellevue Medical Center
Stice, Glen, Educational Testing Service
Stoker, Howard W., Florida State University
Stoney, S. David, South Carolina State Department of Education
Stoughton, Robert W., Connecticut State Department of Education
Stuart, William A., Educational Testing Service
Sussman, Leah, Newark (New Jersey) Public School System
Swanson, Edward O., University of Minnesota
Sweeney, Edward J., Life Insurance Agency Management Association
Swineford, Frances, Educational Testing Service
Symonds, J. P., New York City
Symonds, Percival M., New York City
Taber, Victor, New York State Department of Education
Tansey, Gertrude, Lakewood (New Jersey) Senior High School
Taylor, Calvin W., University of Utah
Taylor, Justine, Educational Testing Service
Taylor, Samuel J., Philadelphia Department of Personnel
Terral, J. E., Educational Testing Service
Thalheimer, Astrid H., Boston University
Thomas, William F., University of Wisconsin
Thorndike, Robert L., Teachers College, Columbia University
Thurstone, Thelma Gwinn, University of North Carolina
Tiedeman, David V., Harvard University

TIFT, Mrs. Floyd, Hillcrest Children's Center, Washington, D. C.

TOOKS, Joyce, Philadelphia Department of Personnel

TRACEY, Ellen K., Oceanside Senior High School, Brooklyn, New York

TRAIL, Stanley M., Rhode Island College of Education

TRAVERS, R., University of Utah

TRAXLER, Arthur E., Educational Records Bureau

TRIGGS, Frances, Committee on Diagnostic Reading Tests, Inc., New York City

TUCKER, Ledyard R, Educational Testing Service

TULLY, G. Emerson, Florida State Univ.

TURNBULL, William W., Educational Testing Service

TURNER, Charles J., Richmond (Virginia) Public Schools

TURNER, Nura D., State University of New York

UPSHAW, Harry S., University of North Carolina

URQUHART, Helen L., Brown University

VALLEY, John R., Educational Testing Service

VAN DAFF, Gracia, U. of North Carolina

VICKERY, K. N., Clemson College, Clemson, South Carolina

WAGNER, E. Paul, State Teachers College, Bloomsburg, Pennsylvania

WAHLGREN, Hardy L., State Teachers College, Geneseo, New York

WALLACE, Wimburn L., The Psychological Corporation

WALLER, Raymond L., Allentown (Pennsylvania) Public Schools

WALLMARK, Madeline, Educational Testing Service

WALSH, John J., Boston College

WALTON, Wesley W., Educational Testing Service

WAMACKS, Cliff C., Temple University

WAMACKS, Lois B., General Electric

WARD, Annie W., University of Tennessee

WARD, Lewis B., Educational Testing Service

WATKINS, Betty Jane, Educational Testing Service

WATKINS, Richard W., Educational Testing Service

WATSON, W. S., The Cooper Union, New York City

WEBB, Sam C., Emory University

WEISS, Eleanor S., Educational Testing Service

WEISS, Joseph, Polytechnic Institute of Brooklyn

WEITZMAN, Ronald, Educational Testing Service

WELLCK, A. A., University of New Mexico

WESMAN, A. G., The Psychological Corporation

WETTEROTH, William J., New York City Police Academy

WHIGHAM, E. L., Wilmington (Delaware) Board of Public Education

WHITLA, Dean K., Harvard College

WILKE, Marguerite M., Greenwich (Connecticut) Public Schools

WILKE, Walter H., New York University

WILLEMIN, Louis P., Dept. of the Army

WILLEY, Clarence F., Norwich University, Northfield, Vermont

WILLIAMS, Clarence M., University of Rochester

WILLIAMS, Roger K., Morgan State College, Baltimore, Maryland

WILSON, Kenneth M., Southern Regional Education Board, Atlanta, Georgia

WINANS, S. David, New Jersey State Department of Education

WINGO, Alfred L., Virginia State Department of Education

WINTERBOTTOM, John, Educational Testing Service

WISNER, Roscoe W., Philadelphia Department of Personnel

WOMER, Frank B., University of Michigan

WRIGHT, Wilbur H., State Teachers College, Geneseo, New York

WRIGHTSTONE, J. Wayne, New York City Board of Education

ZEIGLER, Martin L., Pennsylvania State University

ZUBIN, Joseph, Columbia University

ZUCKERMAN, Harold, New York City Board of Education